Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biomedical Laboratory Science

Yuna Shin

Identification of NRF2 binding sites in metastatic breast cancer cells by ChIP-seq

Master's thesis in Molecular Medicine
Supervisor: Kristin Gabestad Nørsett
Co-supervisor: Geir Bjørkøy
June 2021

NTNU
Norwegian University of
Science and Technology

Yuna Shin

Identification of NRF2 binding sites in metastatic breast cancer cells by ChIP-seq

Master's thesis in Molecular Medicine
Supervisor: Kristin Gabestad Nørsett
Co-supervisor: Geir Bjørkøy
June 2021

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biomedical Laboratory Science

NTNU
Norwegian University of
Science and Technology

Abstract

Background: Nuclear factor erythroid 2-related factor 2 (NRF2) is a transcription factor that is indicated to be an important component in promoting cancer metastasis. Accordingly, the role of the transcription factor has been investigated in diverse types of aggressive cancer. However, a comprehensive identification of NRF2 functions in metastatic breast cancer is rather limited despite being one of the leading causes of death worldwide.

Objective: This study aims to identify NRF2 binding sites in 66cl4 murine metastatic breast cancer cell line by chromatin immunoprecipitation sequencing (ChIP-seq) to contribute to the understanding of how NRF2 contributes to aggressive breast cancer development.

Results: The genome-scale investigation was conducted with 66cl4 NT cells that constitutively express NRF2 and 66cl4 NRF2 KD cells that are depleted in NRF2. The ChIP procedure was initially optimized to suit the biological conditions of the 66cl4 cell line. The tailored method produced six biological ChIP replicates where five were selected for high throughput sequencing following quality control. Grouped samples were pooled for bioinformatical analyses which enriched 94 significant ChIP-seq peaks in NRF2 antibody precipitated-66cl4 NT samples whereas only one peak was detected in 66cl4 NRF2 KD 3B7. A total of 40 known-NRF2-regulated genes were identified in vicinity of 35 peaks, while the remaining 59 have not been previously identified. Selected peaks were validated using ChIP-qPCR. The results indicate a possible finding of novel NRF2-binding sites in 66cl4 cells.

Discussion: The 94 ChIP-peaks are NRF2-binding sites with both known and novel NRF2-targeted genes located in proximity. This was additionally verified as majority of the peaks were positioned in intronic and intergenic regions which are in line with large-scale transcription factor-binding site studies. Intriguingly, the present study shows that NRF2 are involved in extensive set of functions such as drug metabolism, multi-drug resistance, heme homeostasis and local immune microenvironment through regulating its target genes. Several of the genes were further identified to be linked with metastatic breast cancer. These results extend upon existing evidence to the complex role NRF2 takes in breast cancer.

Acknowledgement

This master's project has been performed at the Autophagy and Oxidative Stress Defense Group of the Centre of Molecular Inflammation Research, Department of Biomedical Laboratory Science, Faculty of Natural Sciences, Norwegian University of Science and Technology as a requirement for the MSc in Molecular Medicine.

First and foremost, I would like to express my special thanks of gratitude to my supervisor Kristin Gabestad Nørsett whose expertise was invaluable in conducting the project and formulating my research topic. Your guidance has pushed me to sharpen my thinking and enhanced my work to a greater level. I also want to thank you for being such an incredible mentor throughout the past year and more. Your positivity, kindness, patience, and encouragement has brought the best out of me. You have made my first research experience ever so enjoyable and inspired me to continue my path in this field.

I would also like to express my sincere gratitude to my co-supervisor Geir Bjørkøy for giving me a golden opportunity to do this amazing project. Working on this NRF2 ChIP-seq project was a wonderful experience. Your insightful feedbacks and unsurpassed knowledge of the topic has been inspirational and provided me with the tools to successfully complete my thesis. I want to thank you for all the support I was given to advance my research.

I would also like to thank Sonja Benedikte Andersen and Kristine Pettersen for your kind guidance around the lab and help with my technical work. You have made my lab experience so much better.

Lastly, a special thanks to my mum, dad, Yunseo, Beli and my amazing friends for always being there for me. I would not have been able to complete this project without you.

Trondheim, June 2021

Yuna Shin

Table of Contents

List of Figures and Table

Abbreviations

ARE	Antioxidant response element
BioCore	Bioinformatics core facility at NTNU
C	Cysteine
Cat.	Catalog number
ChEA	ChIP-X Enrichment Analysis
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CSF1	Colony stimulating factor 1
CUL3	Cullin 3
C_t	Cycle threshold
CXCL3	C-X-C motif chemokine ligand 3
DHS	DNase I hypersensitive sites
dsDNA	Double-stranded DNA
D1Z9C	NRF2 (D1Z9C) rabbit monoclonal antibody (Cell signaling Technology)
EDTA	Ethylenediamine tetraacetic acid
ENCODE	The Encyclopedia of DNA Elements
EP1808Y	Recombinant anti-NRF2 antibody [EP1808Y] (Abcam)
FBS	Fetal Bovine Serum
GCF	Genomics Core Facility at NTNU
GSEA	Gene set enrichment analysis
GST	Glutathione S-transferases
KEAP1	Kelch-like ECH-associated protein 1
KD	Knockdown
Kbp	Kilo-base pairs
IL-1	Interleukin-1
IL-1R2	Interleukin-1 receptor type 2
IL-10	Interleukin-10
IP	Immunoprecipitation
kDa	Kilodalton

Library prep	Library preparation
mAb	Monoclonal antibody
Mbp	Mega-base pairs
MEFs	Mouse embryonic fibroblasts
min	Minutes
mRNA	Messenger ribonucleic acid
Neh	NRF2-ECH homology 2
NFAT5	Nuclear factor of activated T cells 5
NFKB2	Nuclear factor kappa B subunit 2
NGS	Next-generation sequencing
NRF2/NFE2L2	Nuclear factor erythroid 2-related factor 2
NT	66cl4 nontarget
NT-input	Input sample of 66cl4 NT cells
NT-NRF2 ChIP	Anti-NRF2 antibody EP1808Y-enriched 66cl4 NT cells
padj	Adjusted p-value
PBS	Phosphate-buffered saline
PPBP	Pro-platelet basic protein
Q score	Quality score
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RNA-seq	Ribonucleic acid-sequencing
ROS	Reactive oxygen species
sec	Second
sMAF	Small musculoaponeurotic fibrosarcoma
TAM	Tumor-associated macrophages
TBST	Tris buffered saline-with Tween
TF	Transcription factor
TME	Tumor microenvironment
Treg	Regulatory T cell
TSS	Transcription start site
3B4	66cl4 NRF2 KD 3B4
3B7	66cl4 NRF2 KD 3B7

1 Introduction

1.1 Dysregulated NRF2 activity in cancer

Cancer was newly diagnosed in approximately 19.3 million cases and caused almost 10.0 million deaths worldwide in 2020 [1]. Although novel diagnostic and therapeutic measures are continuously advancing, there are still urgent needs for improvement. Cancer is a complex disease which can be initiated by various types of genetic alterations in somatic or germline cells caused by either endogenous or exogenous DNA damaging agents/genotoxic compounds [2, 3]. Cancer is also characterized with tumor heterogeneity which the same type of cancer in different patients consists of a distinct set of mutations which can lead to individual progression and responses against the same treatment [4]. Thus, analyzing cancer genomes in a detailed picture are essential for understanding oncogenic mechanisms and ultimately improving patients' lives.

Ever since the completion of the Human Genome Project, the next-generation sequencing (NGS) technology has improved tremendously and has been applied in cancer-genome sequencing projects worldwide [5]. For instance, the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium has analysed 2,658 whole cancer genomes across 38 types of cancer revealing new mutational discoveries [6]. Various sequencing studies have delivered stratified data on numerous types of mutations and is providing insight on detailed genomic diagnosis, prediction of cancer progression and probability of responding to existing treatments [7].

Some projects have additionally identified dysregulated transcription factor (TF) activity incited by genetic modifications which play a key role in diverse types of cancers [8]. Genetic

alterations in approximately 290 TFs have been found to be involved in oncogenic progression [9]. Among them is the nuclear factor erythroid 2-related factor 2 (NFE2L2), also called NRF2 [10].

NRF2 is a TF encoded by the *NFE2L2* gene which responds against oxidative stress by binding to antioxidant response elements (AREs) and initiates expression of phase II detoxification enzymes [11]. Phase II enzymes mainly have cytoprotective roles against oxidants or toxic compounds by performing conjugating reactions [12]. While it has been found that phase II enzymes play an important role in cancer prevention, there has additionally been findings that disrupted NRF2 can inversely promote cancer growth [13]. Multiple genomic sequencing studies have made progress in revealing the NRF2-mechanisms in various cancers. A project which conducted chromatin immunoprecipitation (ChIP) followed by massive parallel DNA sequencing (ChIP-seq) with non-small-cell lung carcinoma cell line discovered a NRF2-regulated focal adhesion pathway which contributes to metastasis [14]. Another ChIP-seq study reported cytoprotective and metabolic functions of NRF2 in mouse hepatoma cell line [15]. Nevertheless, comprehension of NRF2 pathways in cancer is still unclear. Hence, further investigation on possible genetic alterations of dysregulated NRF2 pathway in cancer is needed.

1.2 The NRF2 pathway

NRF2 is a cap'n'collar basic leucine zipper TF expressed in various tissues and promotes the expression of cytoprotective genes against electrophilic or oxidative stress [16]. However, NRF2 levels are low in healthy conditions due to regulation by Cullin 3 (CUL3) E3 ubiquitin ligase complex and its adaptor subunit Kelch-like ECH-associated protein 1 (KEAP1) in the cytoplasm (**Figure 1. A**) [17, 18]. KEAP1 functions as a redox sensor and a NRF2 repressor while CUL3 is a E3 ubiquitin ligase which together mediates the NRF2 ubiquitin-proteasome

pathway [19, 20]. In these basal states, NRF2 is bound to the Kelch domain of KEAP1 homodimer in a “Hinge and Latch” method with its NRF2-ECH homology 2 (Neh2) domain which consists of two binding sites referred as ETGE (hinge) and DLG (latch) motifs with each distinctive binding affinities (high and low, respectively) [21]. Interaction via both motifs ensures tight regulation of NRF2 by KEAP1. Subsequently, seven lysine residues located between the two motifs in NRF2 are targeted by KEAP1-CUL3 complex for poly-ubiquitination ultimately leading to degradation by 26S proteosomes (**Figure 1. A**) [17]. Other than KEAP1-CUL3 mediated pathways, there are additional complexes such as β-transducin repeat-containing protein (β-TrCP)-CUL1 which also participates in the NRF2 degradation pathway. However, KEAP1-CUL3 is known to be the predominant proteins of the process whereas β-TrCP-CUL1 is supplementary [22].

On the contrary, once the cell is exposed to reactive oxygen species (ROS) or stress-inducing chemicals, NRF2 dissociates from the KEAP1-CUL3 complex (**Figure 1. B**) [23]. This occur in different ways. Cells can respond to oxidative stress by oxidation of thiol groups of cysteine residues in KEAP1 such as C (Cysteine)151, C273, C288 and release NRF2 through detaching the latch DLG motif from KEAP1. KEAP1 is then degraded by selective autophagy through binding to p62 or sequestered into inactive complexes by the same protein [21, 24-26]. Separating CUL3 from KEAP1 also contributes to disjoining NRF2 from the composition. Electrophile-induced modification of C151 in KEAP1 was found to directly trigger dissociation of CUL3 from the KEAP1-CUL3 complex freeing NRF2 from ubiquitylation [27]. Both pathways stabilize and release NRF2 from the complex, allowing it to accumulate and translocate to the nucleus (**Figure 1. B**) [28].

In the nucleus, small musculoaponeurotic fibrosarcoma (sMAF), a transcriptional activator, heterodimerizes to the Neh1 domain of NRF2 and together binds to ARE, a regulator of redox

homeostasis (**Figure 1. B**) [20, 26]. It has also been discovered that the transcriptional coactivator p300/CBP can acetylate the Neh1 domain of NRF2, augmenting the binding to DNA and commencing antioxidant response [29]. MED16, a mediator complex, then binds to NRF2 and functions as an intermediate to transfer transcription activating signals to the proteins at the promoter site [26, 30]. This process initiates expression of downstream antioxidant or detoxification genes. In contrast, NRF2 also functions as a transcription inhibitor of proinflammatory cytokines such as IL-1β by interfering in the binding of RNA polymerase II to their transcription start sites (TSSs) in a method independent from ROS levels [22]. Once the redox balance is reached, KEAP1 promotes the translocation of NRF2 from the nucleus to the cytoplasm by interacting with KPNA6 (Karyopherin α6). NRF2 is thereupon proteolyzed by the ubiquitin-proteasome system which decreases NRF2 back to the basal level [31, 32].

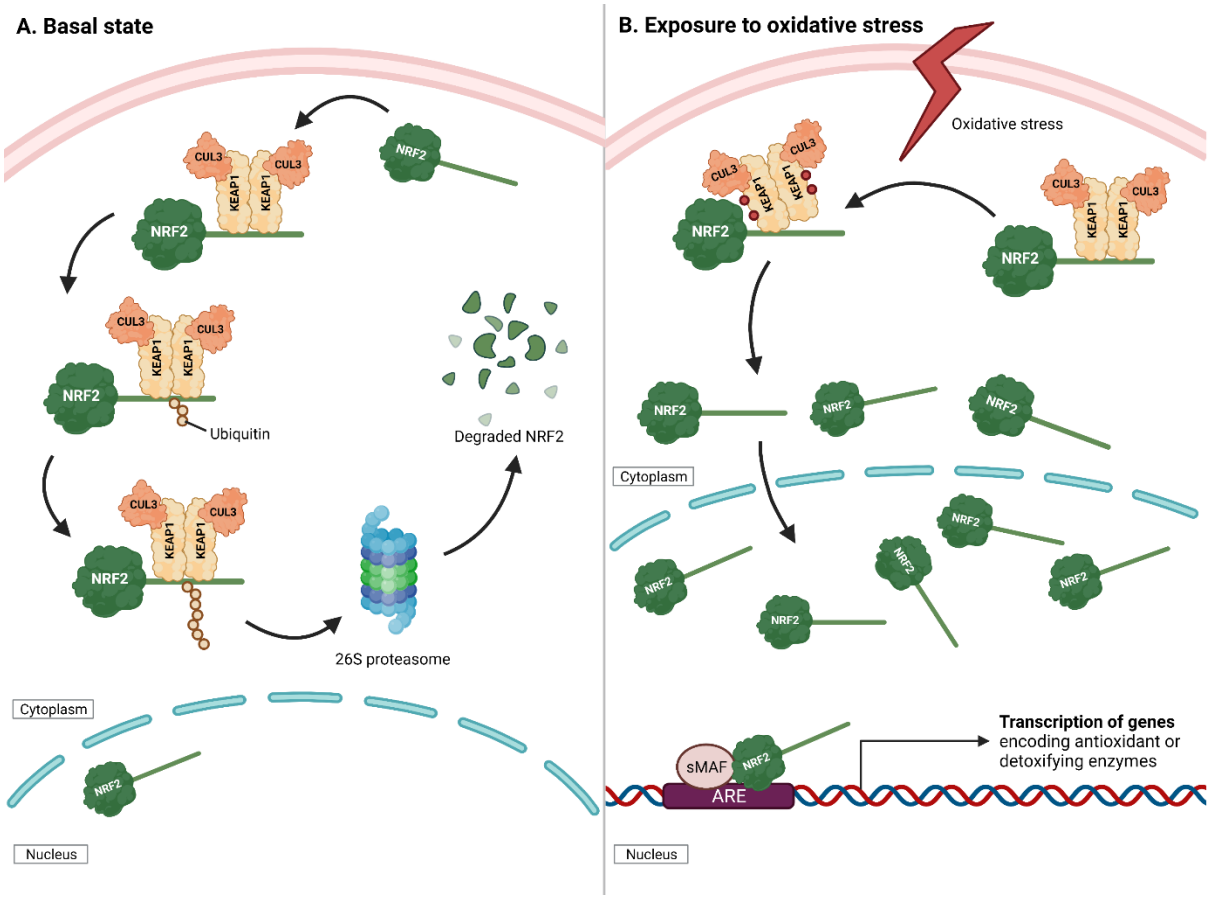


Figure 1. When a cell is exposed to oxidative stress, NRF2 rapidly accumulates and induces transcription of cytoprotective genes. (A) In basal states, NRF2 concentration is maintained at a low level and only a small portion are active. KEAP1-CUL3 complex polyubiquitinates NRF2 which subsequently undergoes proteasomal degradation. **(B)** However, NRF2 manages to escape the KEAP1-CUL3 complex into the nucleus and actively induces expression of antioxidant and detoxifying genes once the cell senses reactive oxygen species and electrophiles. Created with BioRender.com.

1.3 Two sides of NRF2

1.3.1 Preventing tumorigenesis

Ever since NRF2 was first isolated [10], the protein has become a focus of studies due to being known as a protector of cells from ROS or xenobiotics and thus its potential to become a key in therapeutic development. Excessive production of ROS can impair cell components and even

cause DNA damage which could ultimately lead to cardiovascular diseases, neurological disorders or carcinogenesis (**Figure 2. A**) [33]. Therefore, activating NRF2 in these circumstances could lead to a better prognosis in patients. The significance of NRF2 was further emphasized in studies where NRF2 depletion induced susceptibility to autoimmune diseases or liver, gastric, prostate, bladder, colon, and breast cancer [34-42]. It was thus clear that NRF2 shielded healthy cells from these dangers by promoting transcription of cytoprotective genes and maintaining redox homeostasis (**Figure 2. A**) [42, 43].

1.3.2 Promoting cancer

On the other hand, contradicting discoveries were later made when constitutively active NRF2 were found in the nucleus of human lung cancer cells due to a somatic mutation in KEAP1 [44]. Since then, many research groups have found similar findings in different tumor types, which suggested that accumulated NRF2 in cancer cells may in fact be a driver for its proliferation, metastasis, and chemo- or radio-resistance through constant expression of cytoprotective genes (**Figure 2. B**). In other words, cancer cells could hijack the properties of NRF2-induced antioxidant response for its own accelerated progression without being halted by therapeutic measures and thus leading to poor prognosis [45].

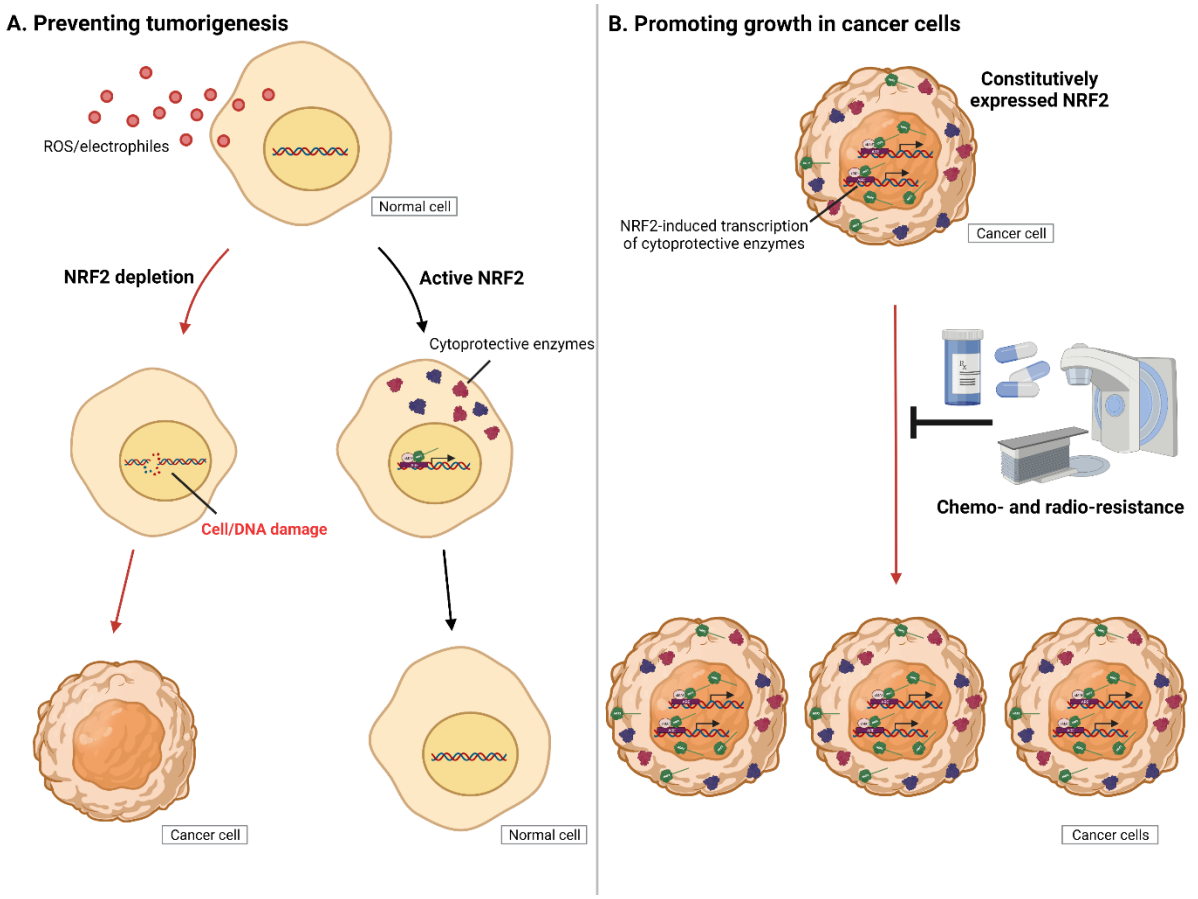


Figure 2. NRF2 plays a contradicting role of protecting healthy cells from tumorigenesis whilst supporting aggressive proliferation and metastasis in cancer cells. (A) When a healthy cell is exposed to oxidative stress, NRF2 facilitates the production of cytoprotective enzymes to prevent cell or DNA impairment. This was proven in studies where cells deficient in NRF2 was discovered to be vulnerable in developing diseases such as cancer. **(B)** Conversely, continuous expression of NRF2 in cancer cells is known to promote cancer progression by hijacking NRF2 properties by resisting against anticancer treatments and reprogramming metabolic pathways. Created with BioRender.com.

1.4 Causes of NRF2 overactivation in cancer

Extensive research has been conducted discovering the molecular mechanisms that leads to constitutively active NRF2 in cancer cells and tissues. Most aberrations were found within the KEAP1-CUL3/NRF2 interaction processes, while mutations in proto-oncogenes or intervention of proteins that regularly do not participate in the NRF2 pathway was also identified to be able to provoke nuclear concentration of NRF2.

1.4.1 Genetic alterations found in KEAP1/NRF2 pathway

Mutations in *NFE2L2*, NRF2 encoding gene, is discovered to induce elevation of the NRF2-ARE pathway in lung, head, neck esophagus and skin squamous cell carcinomas (SCC) [46-48]. One of the mutations are exon 2 deletions. Exon 2 encodes the Neh 2 domain encompassing the KEAP1-binding DLG and ETGE motif and its deletion was detected in both mRNA and protein levels in the cancer types [48]. Missense mutations within or near the DLG and ETGE domain also contributed to substantial amounts of NRF2 accumulation [47]. This suggests that the deletions or missense mutations in *NFE2L2* can impair the interaction of NRF2 with KEAP1 and thus stabilize NRF2 activity.

Low expression of KEAP1 due to either mutations or aberrant epigenetic mechanisms was also found to be a cause of abnormal NRF2 signaling in tumors such as SCC, lung, gastric, hepatocellular, and breast cancer [46, 49, 50]. Missense mutations, causing an allelic loss of *KEAP1* locus, upregulated NRF2 signaling and antioxidant gene expression which supports cancer cell survival [50]. Furthermore, Wang et al. suggested hypermethylation of *KEAP1* gene as another contribution to suppressed KEAP1 activity. The group found hypermethylation in promoter regions of *KEAP1* in human cancer cell lines and tissues. Three highly methylated CpG sites were found and recognized to be binding sites of *KEAP1* transcription regulators and thus being indicated as a major cause of constitutive NRF2 activation [49].

Overall, modifications of *KEAP1* were mainly characterized with loss-of-function, whereas those in *NFE2L2* led to a gain-of-function [46]. Despite the variety of mutation types in *KEAP1*, and *NFE2L2*, they all result in increased NRF2 concentration and antioxidative response. However, mutations in *KEAP1* seem to be mutually exclusive with mutations in *NFE2L2* [46, 51, 52]. The Cancer Genome Atlas Research network noticed mutually exclusive mutation patterns of *NFE2L2* with *KEAP1* [46], whilst Leiserson et al. further found a significant mutual

exclusivity of mutations in *KEAP1*, *NFE2L2*, and other proteins within the NRF2 pathway through their pan-cancer network analysis [51]. This is supported by the observation that several genetic alterations typically do not happen within the same pathway as a second mutation will not provide selective advantage to the pre-existing mutation within a cancer cell [53]. Mutual exclusion of mutations between *NFE2L2* and *KEAP1* are in line with the findings that modifications of the pathway converge towards to the same result of NRF2 stabilization.

1.4.2 NRF2 upregulation provoked by external factors

NRF2 expression can be increased by mutations in proto-oncogenes such as *c-Myc*, *K-Ras*, *B-Raf*. These oncogenes induce constitutive transcription of *NFE2L2* by directing Jun and Myc, oncogenic TFs, to directly bind with the TSS of *NFE2L2*. Increased NRF2 expression was further revealed to initiate lung and pancreatic tumorigenesis [18].

Moreover, NRF2 production can be amplified by factors which are able to interfere in NRF2/KEAP1 pathways such as p62/SQSTM1 (hereafter p62) and PALB2. p62 is a protein of autophagic pathway and functions as a cargo receptor for cellular wastes that are to be degraded by lysosomes [54]. Since p62 is degraded with its cargo in the lysosomes, defective autophagy results in accumulation of p62 within the cell. Overproduction of p62 promotes binding of the receptors to KEAP1, sequestering the protein into inclusions. This releases NRF2 and cytoprotective enzymes are constitutively expressed, consequently eliciting pathological conditions [55].

PALB2 was also found to compete with NRF2 for binding with KEAP1 and induce elevated NRF2 activity. PALB2 is a binding partner and localizer of BRCA2 and functions in breast cancer suppression. The protein contains an ETGE motif which directly binds with KEAP1, therefore inducing high NRF2-ARE association when PALB2 is overexpressed. Upregulated

production of PALB2 has been found in breast, lung, colon and pancreas cancers, strengthening the postulation of PALB2 promoting the NRF2 activity [56].

1.5 NRF2 in breast cancer

NRF2 is recognized as a significant factor of cancer development due to its function of upregulating proteins participating in metabolic pathways and how this promotes carcinogenesis once the basal mechanisms are aberrated [57]. However, majority of NRF2-cancer correlations have been discovered in either non-small-cell lung cancers or SCCs, whereas the knowledge in breast cancer is rather limited.

1.5.1 Metastatic breast cancer

Breast cancer is one of the most diagnosed cancer which was recorded with the highest number of new diagnosis (2.26 million cases) in 2020 and leads second place in cause of death among women worldwide [58]. There have been major improvements in research which has led to a steady increase of survival rates over the past few decades. Even so, therapeutic methods for metastatic breast cancers are insubstantial compared to primary tumors diagnosed at an early stage [59]. Metastatic types take up most of the breast cancer death rates. For patients with metastatic breast cancer, the estimated 5-year survival rate since its diagnosis is found to be 28% for women, and 22% for men [60]. Thus, identifying molecular causes that supports metastasis of breast cancer is crucial. Although several studies have found NRF2 upregulation to promote metastatic properties and sustain its malignancy in breast cancer cell lines [28, 61], a detailed system of how NRF2-regulated genes contribute to this remains unknown.

1.5.2 Breast cancer recurrence

Breast cancer recurrence is another challenge as relapsed cancer contain large amounts of aggressive cells which is connected to an increased mortality rate [59]. Predicting breast cancer relapse is complex since it is dependent on a variety of factors such as the subtype of breast cancer, length of dormancy periods and social or economic backgrounds of the patient [59, 62]. Therefore, finding a concrete molecular prognostic method for recurrence is vital.

Several studies have likewise mentioned the importance of developing prognostic signatures of recurring breast cancer and further stated the significant role NRF2 takes in the process [61, 63]. Fox et al. has recently suggested that constitutively active NRF2 is a critical component in augmenting breast cancer relapse irrespective of the subtype. According to the study, when HER2 is therapeutically inhibited, ROS levels are increased due to metabolic changes within the cell. Although escalated ROS levels induce cell death, few cancer cells survive and remain dormant until it recurs. The dormant breast cancer cells were found to upregulate NRF2 in response against the ROS state. Subsequently, elevated NRF2 signaling regulates redox homeostasis and nucleotide synthesis which together promotes breast cancer recurrence [63].

There are additional findings which indicate the contribution of NRF2 to aggressive breast cancer development [61-65]. However, further explanations in the underlying mechanisms of NRF2 and its target gene regulations in malignant breast tumors are required.

1.6 Selecting ChIP-seq as a method for identifying NRF2-binding sites

Discovering TF-binding sites in DNA have been of great interest for decades which has led to the advancement of experimental methods for its identification. There are traditional methods such as electrophoresis mobility shift (EMSA) assays and DNase I hypersensitive sites

sequencing (DNaseI-seq). Meanwhile there are ChIP-based methods, namely ChIP integrated with DNA microarray (ChIP-chip) and ChIP-seq [66]. Although each method has been well-used in studies, most also have presented a few drawbacks. Nucleotide and protein complexes often dissociate in EMSA which hinders detection of the complex, while DNaseI-seq has limitations in differentiating similar protein-binding sites [67]. ChIP-chip enables a global analysis of DNA-TF interactions but was characterized with low resolution due to its dependency on annealing to immobilized oligonucleotides on microarray tiles which are oftentimes ineffective [67, 68]. However, ChIP-seq has largely exhibited advantages over other approaches in producing accurate reads of DNA sequences [69].

ChIP-seq is a variation of a ChIP assay integrated with massive parallel sequencing that is used to map genomic binding locations of TFs in living cells by enriching the DNA binding sites of interest [67, 70]. It includes techniques of crosslinking protein-DNA complexes, fragmenting the DNA, immunoprecipitating genes of interest, reverse-crosslinking complexes and sequencing the precipitated DNA [71]. This procedure is marked by the ability to read millions of DNA fragments with high efficiency and resolution. It detects the interactions between proteins and DNA, thus providing information on gene regulation events [67]. It additionally offers high and even coverage reads of sequences compared to other sequencing methods [72].

In this study, ChIP-seq was chosen as the tool to detect novel NRF2-targeted genes in 66cl4 metastatic breast cancer cell lines. The experimental procedure had to meet certain requirements for the aim of accurately identifying genes regulated by NRF2. The method first had to allow highly sensitive precipitation for all genes targeted by a TF in mammalian cells *in vitro*. Moreover, it were to encompass high throughput whole-genome sequencing for it to read genes of interest in exons, introns and intergenic regions [73]. Lastly, it had to be a method that can provide an understanding of gene regulation in cancer progression on a genome-wide scale

[74]. ChIP-seq was the method that comprised all abovementioned features and was accordingly selected for this research.

1.7 The 4T1 model of metastatic breast cancer

In the present study, 66cl4 cell lines were used. 66cl4 cells derive from the 4T1 murine mammary tumor model which are isolated from tumors in BALB/cJ mice [75]. The 4T1 model consists of five cell lines with different metastatic propensity including the 66cl4 and 67NR cell lines [75-77]. Although both cell lines form primary tumors in the mammary fat pad of BALB/cJ mice, 67NR cells are unable to metastasize while 66cl4 cells generate secondary tumors in the lung [75]. To recognize the NRF2 functions and its targeted genes in aggressive breast cancer development, the 66cl4 murine breast cancer cell line with metastatic property was selected.

As hyperactivation of the NRF2 is correlated with poor prognosis in various types of cancers, 66cl4 cells that constitutively express the TF was therefore an adequate cell line for NRF2 investigation [28, 56, 61, 63-65]. Furthermore, two NRF2 knockdown (KD) 66cl4 cell lines were used to differentiate the binding sites of NRF2 to the 66cl4 nontarget (NT) cell line. NRF2 KD cells can be characterized by its susceptibility to ROS-induced cell damage as NRF2 depleted mice are shown to be more sensitive to ROS and exogenous chemicals due to the absence of NRF2-induced antioxidant enzymes [78, 79].

1.8 Objective of this study

The overall aim of this study was to identify NRF2 binding sites in 66cl4 metastatic breast cancer cell lines by ChIP-seq.

Previous studies from our research group compared gene expression differences between 66cl4 cells and the non-metastatic 67NR cells and reported that constitutive activation of NRF2 could promote metastasis of breast cancer [61]. The findings strengthened prior studies addressing the significance of NRF2 in malignant breast tumors [28, 61, 63-65]. However, comprehensive identification of NRF2-targeted genes in metastatic breast cancer cells remains limited and requires further research.

The specific objectives were:

- Optimize protocols of chromatin sonication cycles, ChIP, and ChIP-qPCR
- Perform NRF2-chromatin crosslinking, sonication, and ChIP of the three 66cl4 cell lines
- ChIP-qPCR with known NRF2-regulated genes for ChIP sample quality assessment
- ChIP-seq
- Comparing list of significantly downregulated genes in 66cl4 NRF2 KD cell lines (RNA-seq) to list of NRF2 targeted genes (ChIP-seq).
- Bioinformatic analyses (Enrichr, cBioPortal) of RNA-seq and ChIP-seq results
- Verification of ChIP-seq results using ChIP-qPCR

2 Materials and Methods

2.1 Experimental Cell lines

As a model for metastatic breast cancer, murine model 66cl4 was chosen. 66cl4 cells were obtained from Barbara Ann Karmanos Cancer Institute. From our research group, ShRNA-NRF2 KD were previously produced by viral transduction (SHC216V-1EA, Cat.: TRCN0000054658, Sigma-Aldrich, MO, USA). In a medium with hexadimethrine bromide (8 µl/ml), the 66cl4 cells were infected with lentiviral particles (MOI 0.1) after 6 hours since seeding. Cells were then split in 1:17, 24 hours after infection. Following 48 hours, puromycin (3.25 µl/ml) was used for 1 week for cell selection. Single cell colonies were collected with cloning cylinders and examined for NRF2 expression [61]. Two KD cell lines with a 90% decrease in expression of NRF2 in both mRNA and protein levels were chosen and named 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7 [61].

2.2 Cell Cultures

The NRF2 KD cell lines and the control group, 66cl4 nontarget (NT), were cultured in DMEM (Cat.: BE12-604 F, Lonza, Switzerland) supplemented with 10% FBS (fetal bovine serum, Cat.: 10270-106, Thermo Fisher Scientific, MA, USA), 1% L-glutamine (Cat.: De-17-605E, Lonza) and 1% Penicillin-streptomycin (Cat.: 15070-063, Thermo Fisher Scientific) as additives in 75 cm flasks. These cells were grown in an incubator set to 37 °C with 5% CO_2 and were sub-cultured into new flasks once they reached approximately 90 % confluency. The medium was then removed and washed twice with pre-warmed phosphate-buffered saline (PBS). Subsequently, the 66cl4 cells adhered to the bottom of the flasks were dissociated with 0.25%

Trypsin/EDTA (Cat.: 25200056, Invitrogen, CA, USA) and incubated in the same condition as it was cultured for 5 to 10 minutes (min). The detached single cells from 66cl4 NRF2 KD 3B4, 66cl4 NRF2 KD 3B7 and 66cl4 NT were then resuspended in DMEM in the approximate ratio of 1:3, 1:12 and 1:10, respectively. The cells were cultured up to maximum 40 passages and were occasionally examined for possible microbial contamination by culturing with Penicillin streptomycin-excluded DMEM.

2.3 Immunoblotting

Approximately 798,000 cells were seeded in each well of a 6-well plate for 66cl4 NT and 66cl4 NRF2 KD 3B7, whilst 2,394,000 cells were seeded in each well for 66cl4 NRF2 KD 3B4 plates. The number of cells were previously optimized to achieve 90 to 100% confluency after two days of incubation. Following the two days, medium was removed from the wells and cells were washed in cold PBS on ice. Subsequently, the cells were lysed with 8M urea (Cat.: 1084870500, Merck, Germany), 0.5% (v/v) Triton X-100 (Cat.: T8787, Sigma-Aldrich), 1M DTT (Cat.: 646563, Sigma-Aldrich), 25X Complete® protease inhibitor (Cat.: 1187350001, Roche, Switzerland), phosphatase inhibitor cocktail II (Cat.: P5726, Sigma-Aldrich) and III (Cat.: P0044, Sigma-Aldrich). The adherent cells were scraped off to be translocated to a precooled 50 ml tube and incubated for 20 min on a shaking platform at 4 °C. The incubated tubes were centrifuged at 13,000 rpm for 15 min in 4 °C for protein extraction. Supernatants were thereafter obtained and aliquoted into six 1.5 ml tubes. Protein concentrations of each sample were measured twice using Bio-Rad protein assay (Cat.: 500-0006, Bio-Rad, CA, USA). Protein extracts were diluted with Bio-Rad Protein Assay Dye Reagent Concentrate (Cat.: 5000006, 1:5000, Bio-Rad) to measure OD_{595} of each sample. OD_{595} measurements were converted to concentration units.

$$\frac{OD_{595} * 22 * dilution\ factor}{1000} = Protein\ Concentration\ (\mu g/\mu l)$$

The selected samples were diluted in Tris-HCl, 4X LDS buffer, and 1M DTT and made sure protein concentrations of each sample were 2.5 µg/µl. The Odyssey One-Color Protein Molecular Weight marker (Cat.: 928-40000, LI-COR Biosciences, NE, USA) was used to monitor the transfer efficiency from gel to blotting membrane and to estimate the size of proteins on the blot. Protein samples were thereafter heated at 80 °C for 10 min to enable measurement of the polypeptide length. Denatured proteins were loaded to washed wells of NuPAGE 4 to 12% Bis-Tris protein gel (Cat.: NP0321PK2, Invitrogen). The samples were run in 1X MOPS Running buffer at 200V for 1 hour. Electrophoresed proteins were subsequently dry blotted to nitrocellulose membranes (Cat.: IB23001, Invitrogen) with iBlot 2 Dry Blotting System (Cat.: IB21001, Invitrogen). To block and decrease background variability, the dry blotted membrane was incubated in Intercept® (TBS) Blocking Buffer (Cat.: 927-60001, LI-COR Biosciences) and 1X tris-buffered saline with tween (TBST) (20 mM Tris-base (Cat.: 10708976001, Sigma-Aldrich), 137 mM NaCl, 0.1% Tween 20 (Cat.: P1379, Sigma-Aldrich)) in dilution of 1:1 in room temperature for 60 min. This was followed by probing the membrane with NRF2 (D1Z9C) XP Rabbit monoclonal antibody (mAb) (Cat.: 12721, 1:1000, Cell Signaling technology, MA, USA) and anti-COX IV antibody (Cat.: ab33985, 1:1000, Abcam, UK) diluted in the Intercept TBS blocking buffer/TBST solution overnight at 4 °C. The membrane was washed with TBST and incubated for 10 min, and the process was repeated twice more. The proteins were detected by applying near-infrared fluorescent (IRDye) secondary antibodies (Cat.: 926-32213, 1:5000, Cat.: 926-68072, 1:5000, LI-COR Biosciences) and incubated in room temperature for 1 hour. Following the washing of membrane in TBS, the fluorescence was visualized with Odyssey Near Infrared scanner (Cat.: CLx, LI-COR Biosciences).

2.4 Chromatin Immunoprecipitation (ChIP)

2.4.1 Cell Plating

Three 75 cm flasks with 90% confluency from each of the 66cl4 cell line were used for plating (**Figure 3. A**). Following medium removal, the cells were washed twice with PBS. The cells were then trypsinized in 0.25% trypsin/EDTA and incubated at 37 °C with 5% CO_2 up to 10 min for proper singularization. Subsequently, 8 ml of DMEM with additives were added to each flask in order to block further trypsinization. Following thorough resuspension of the cell aggregates, the cells underwent assessment with a microscope to confirm cell singularization. The singular cells were collected in one flask for each cell line.

Prior to distributing cells to sterile 150 cm dishes, cell numbers were counted with an automated cell counter EVE™ (Cat.: 6284817, Buch & Holm, Denmark) to ensure accuracy. This was done by collecting 10 µl of cells from flasks to two separate 1.5 ml reaction tubes (Cat.: 72.690.001, Sarstedt, NC, USA) and staining with 10 µl of tryptophan blue. After pipetting, 10 µl were removed and inserted to both sides of the slide for calculation. The mean number of live cells from the slide was then used to evaluate the volume needed to be seeded to each plate. The number of dishes depended on the number of antibodies to be used with an additional plate for cell counting.

Including the resuspended cell volume and complete medium (DMEM, 10% FBS, 1% L-glutamine, 1% P/S), a total of 20 ml was plated to each of the four dishes per cell line (**Figure 3. A**). These cells were grown in the 37 °C and 5% CO_2 incubator for approximately 48 hours.

2.4.2 Crosslinking and Harvesting

Incubated cells were washed twice with pre-warmed PBS to remove the remaining complete medium. This was followed by adding more warmed PBS to each dish. For crosslinking of proteins and DNA [80], cells were treated with 0.5 % formaldehyde solution (Cat.: 100496, Merck) and incubated on rocking platform shakers for 10 min in room temperature (**Figure 3. B**). Formaldehyde was then quenched by addition of 2 ml of 1.25M glycine which was continued with shaking incubation for 5 min. Crosslinking buffers were then removed, and cells were washed twice with cold PBS. To harvest cells, 6 ml of cell detaching buffer (PBS, 1mM EDTA, 1X Complete) were applied to the dish. By scraping the dish, loosened cells were gathered into a 50 ml tube for centrifugation at 3000 rpm for 5 minat 4 °C. The buffer was poured out after centrifugation and remaining cell pellets were resuspended in 1ml of cell lysis buffer (RIPA/Glycine buffer (10 mM Tris pH 8.0, 1mM EDTA, 140 mM NaCl, 1 % Triton X-100, 0.1% SDS, 0.1 % Na-Deoxycholate, 125 mM Glycine), 1X Complete) for extraction of crosslinked protein-DNA complexes into the solution (**Figure 3. C**). Lysed pellets in the solution were transferred to 1.5 ml reaction tubes and stored at -80 °C.

2.4.3 Sonication

Chromatin lysates from each cell line were thawed and fragmented with Bioruptor® Pico sonication device (Cat.: B01060010, Diagenode, Belgium) (**Figure 3. D**). Thawed samples were resuspended in lysis solution (RIPA buffer, 1X Complete) and transferred to 1.5 ml Bioruptor® Pico microtubes (Cat.: C30010016, Diagenode). Samples were sonicated by 12 cycles with 30 sec ON/30 sec OFF. Sheared chromatin fractions were collected by centrifuging the sonicated tubes at top speed for 15 min at 4 °C. Pellets were resuspended in TE-buffer (10 mM Tris pH 8.0, 1 mM EDTA) to estimate the percentage of solubilized DNA. Input samples

were prepared by taking 10% of the chromatin volume for digestion control and later usage for quantitative polymerase chain reaction (qPCR) and ChIP-seq. Remaining chromatin samples were snap-frozen in liquid nitrogen and stored at -80 °C.

Input and pellet samples were reverse crosslinked by adding 1 µl of RNAseA (10 mg/ml, Cat.: 19101, Qiagen, Germany) and incubating at 37 °C for 30 minu. This was followed by treatment with 5.0 µl of SDS (10%), 2.5 µl Proteinase K (20.6 mg/ml, Cat.: 03115828001, Roche), incubation at 55 °C for 1 hour and subsequently with 65 °C incubation overnight.

Chromatin from input and pellet samples were then purified with QIAquick PCR purification kit (Cat.: 28104, Qiagen). The samples were treated with protein binding buffer (PB) and relocated to QIAquick columns and collection tubes to enable binding of chromatin to spin-column membranes. Following removal of the buffer by centrifuging for 1 min at room temperature, residues on the membranes were washed with wash buffer (PE) and centrifuged twice for complete residue removal. Spin columns were transferred to new 1.5 ml reaction tubes and treated with 50 µl of TE buffer and centrifuged 1 min for DNA elution.

Gathered DNA samples underwent both concentration and fragment length measurement (**Figure 3. E**). DNA concentration was analyzed with NanoDrop 1000 spectrophotometer (Cat.: ND-1000, Thermo Fisher Scientific) by setting up buffer TE as blank. Obtaining the concentration results from both input and pellet samples, volumes corresponding to 500 ng of DNA were applied to each well with 10X loading buffer and TE buffer in 2% agarose gel for electrophoresis. The gel was run for 50 min at 100 V in 1 X TAE buffer and subsequently stained in 0.03% GelRed nucleic acid gel stain (Biotium, CA, USA) solution for 30 min and examined under UV-illumination using Gel Logic (Cat.: 212PRO, Carestream, NY, USA)

2.4.4 Immunoprecipitation (IP)

After verifying the 200-500 bp fragmentation, immunoprecipitation was followed to selectively enrich the NRF2-targeted gene complex (**Figure 3. F**). Protein A magnetic beads (Cat.: 10001D, Dynabeads, Invitrogen), protein G magnetic beads (Cat.: 10003D, Dynabeads, Invitrogen) and selected antibodies known for its high NRF2-affinity were prepared.

A mixture of protein A and G magnetic beads (hereby, protein A/G, Thermo Fisher Scientific) were prepared by adding 250 µl from each into a 1.5 ml DNA LoBind tube (Cat.: 0030108051, Eppendorf, Germany). LiCl wash buffer (250 mM LiCl, 10mM Tris pH 8.0, 1 mM EDTA, 0.5 % NP-40, 0.5 % Na-Deoxycholate) and RIPA buffer (10 mM Tris pH 8.0, 1mM EDTA, 140 mM NaCl, 1 % Triton X-100, 0.1% SDS, 0.1 % Na-Deoxycholate) were also prepared to effectively wash out nonspecific chromatin that may bind to the magnetic beads [81]. Using a magnetic stand (Cat.: 12321D, DynaMag™-2 Magnet, Invitrogen), protein A/G bead mixture was washed twice with 1ml of RIPA buffer containing 0.5X complete to draw out potential nonspecific components from the mixture.

Thawed chromatin samples from each cell line were divided into three different 1.5 ml reaction tubes corresponding to the types of antibodies to be used. The chromatin volume in each tube depended on the amount needed for each antibody. The samples in each tube were precleared with protein A/G bead mixture to remove potentially reactive nonspecific lysate components and was incubated at 4 °C for 1 hour with a rotating tube mixer [82].

Supernatants separated from the beads were transferred to DNA LoBind tubes. Precleared chromatin was treated with two control antibodies and two NRF2-targeting antibodies (**Figure 3. F-a**). Normal rabbit Ig G (Cat.: 2729S, Cell Signaling technology) was used for negative control as it does not bind to any known human antigen [83], whereas Histone H3 rabbit mAb

(Cat.: 4620, Cell Signaling technology) was used for positive control due to its specificity for histone H3, a core protein that is bound to most sequences [84]. For NRF2 mAbs, antibody D1Z9C and recombinant anti-NRF2 antibody (EP1808Y) (Cat.: ab62352, Abcam) were both initially tested in the first two sets of samples to compare effectivity in immunoprecipitating NRF2-bound target genes in 66cl4 cell lines. Following the manufacturer's recommendations, the negative control, Normal Rabbit IgG, was added by 1 µg to 1 µg of chromatin while 5 µl of the positive control, histone H3 mAb, was added to 5 µg of chromatin [83, 85]. For anti-NRF2 mAbs, 2.0 µg of EP1808Y was added to 25 µg of chromatin while 5 µl of D1Z9C was applied to 10 µg of chromatin [80, 86]. The ChIP samples were then incubated in rotation at 4 °C overnight.

The procedure continued with the application of protein A/G beads for affinity purification (**Figure 3. F-b**). This was incubated for 2-3 hours at 4 °C in rotation for binding of the complex to Fc regions of antibodies. The beads were subsequently washed with RIPA wash buffer (RIPA buffer, 0.5 X Complete) five times, LiCl wash buffer (LiCl buffer, 0.5 X Complete) once and TE buffer once each for 5 min in rotating incubation at 4 °C (**Figure 3. F-c**). Washing buffers were removed with magnetic stand and beads were resuspended in TE buffer containing 1 % v/v RNaseA which was followed with incubation at 37 °C for 30 min in shaking of 1200 rpm for consistent bead suspension. The sample was then treated with 0.46 % SDS and 0.05 % Proteinase K to remove proteins from the linked DNA and prevent its degradation from nucleases (**Figure 3. G**). This was incubated in same shaking speed at 55 °C for 1 hour. The incubation temperature was increased to 65 °C and left for shaking overnight.

Reverse crosslinked samples were purified with QIAquick PCR purification kit (**Figure 3. H**). For spin-column binding, aggregated beads were removed on magnetic stand after buffer PB was added. Suspended samples were transferred to spin-columns and centrifuged for 1 min at

room temperature. Chromatin bound to the membrane was washed with buffer PE and centrifuged twice in the same condition for thorough removal of residues and buffer. The columns were placed in new tubes and DNA was eluted from the membrane with 30 µl of TE buffer and 1 min of centrifugation. Purified samples were stored at -20 °C.

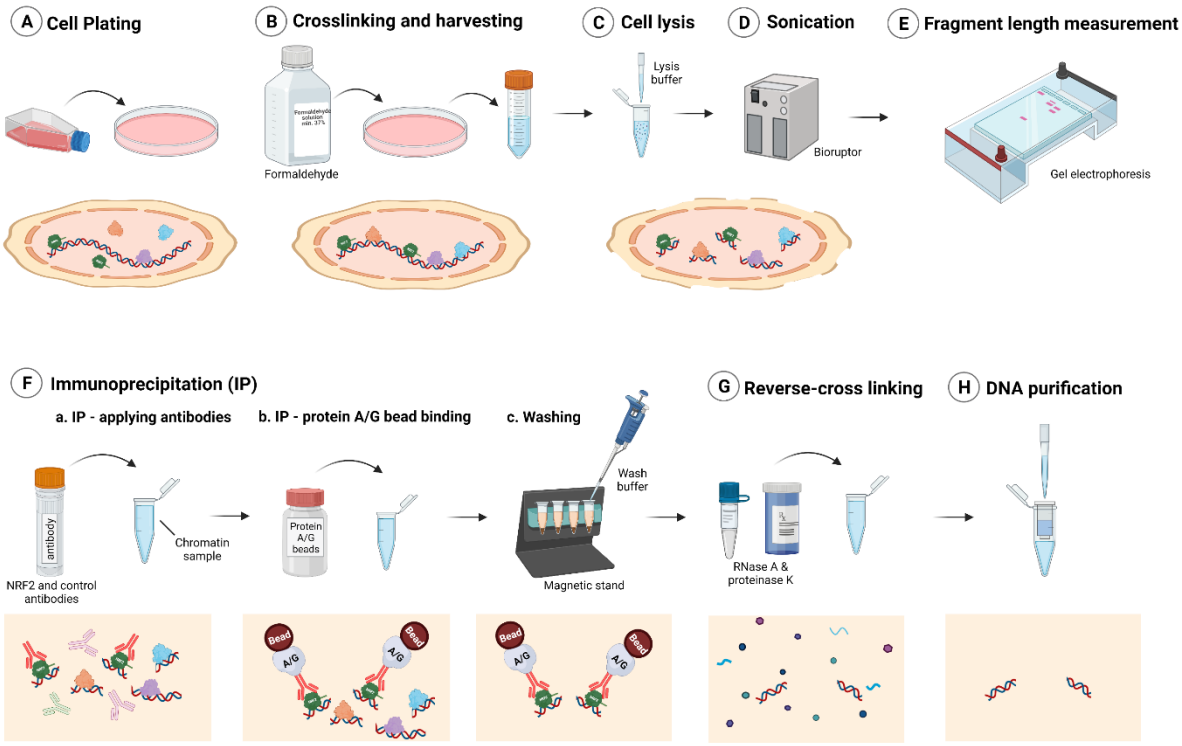


Figure 3. Summarized ChIP procedure. The chromatin immunoprecipitation (ChIP) process executed in this study mainly consisted of eight steps with a goal of extracting the DNA fragment of interest. Each experimental stage is illustrated along with a diagram below which visualizes the molecular state occurring in each stage. Created with BioRender.com.

2.5 ChIP-qPCR

ChIP samples were diluted in 1:2 of TE buffer. For amplification, each primer's master mix was prepared with 4 µl of RNase-free water, 10 µl of SYBR Green master mix (Cat.: 15350929, Applied Biosystems, CA, USA) and 2 µl of each forward and reverse primer solutions (2.5 µM) per well.

For ChIP-qPCR that was conducted for examining the quality of ChIP samples prior to sequencing, SimpleChIP® Mouse RPL30 Intron 2 primer (Cat.: 7051S, Cell Signaling Technology) was selected as a positive control while primers of putative NRF2-regulated genes were referenced from a NRF2 ChIP-seq study by Malhotra et al. and obtained from Sigma-Aldrich (**Supplementary Table 1**) [87, 88]. On the other hand, for ChIP-qPCR conducted for validating ChIP-sequenced datasets, primers were designed in the present study using ChIP-seq peaks. With the peak summit as the midpoint, DNA sequences of approximately 150 to 200 bp were selected from each of nine significant peaks. These peaks were located within or proximate to TSSs of known NRF2-associated genes. The sequences were thereafter tailored with an online PCR primer design tool called GenScript (https://www.genscript.com/tools/pcr-primers-designer) to generate primers for both strands with an approximate length of 90 bp (**Supplementary Table 2**).

The 18 µl of primer containing master mixes were allocated into 2 µl of sample DNA for one well. Each sample was prepared as triplicates in a 96-well reaction plate (Cat.: N8010560, Applied Biosystems). All real-time qPCR was performed by StepOne Real-Time PCR system (Cat.: 4376357, Applied Biosystems). The reaction process was initiated by first heating up to 50 °C for 2 min and subsequently 95 °C for another 2 min for double-stranded DNA (dsDNA) denaturation. Amplification cycling was then run for 40 cycles of 3 seconds (sec) at 95°C for

denaturing and 30 sec at 60°C for primer annealing and extension. Melting curve analysis followed starting with 15 sec at 95°C, 1 min at 60°C and steadily increasing by 0.3°C to 15 sec of 95°C to confirm amplification specificity.

To differentiate the amount of amplified DNA between different cell lines and IP samples, percent input method was used. This was performed by normalizing the cycle threshold (C_t) signals of IP results with signals from input sample results. The purpose was to exclude possible quantification inaccuracies resulting from variability within samples from factors such as IP efficiency and DNA recovery. The input was first adjusted by subtracting $\log_2 10$ from the C_t value of inputs, as the chromatin volume in input samples were only 10 % of the chromatin and thus have a dilution factor of 10. The adjusted input was thereupon subtracted by C_t of IP samples which equates ΔC_t. To retrieve the percent input of each sample for comparison in relative gene quantity, -ΔC_t was powered by 2 and multiplied by 100 [89].

$$Adjusted\ Input\ C_t = Input\ C_t - \log_2 10$$

$$\Delta C_t = Adjusted\ Input\ C_t - IP\ C_t$$

$$Percent\ Input = 100 \times 2^{-\Delta C_t}$$

2.6 RNA-sequencing

RNA sequencing (hereby, RNA-seq) data, produced by Bjørkøy et al. (unpublished), was analyzed in the present study. Bjørkøy et al. sequenced RNA samples from the three identical cell lines as those used in this study which were 66cl4 NT, 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7. To identify RNA sequences that are differentially expressed in NRF2 KDs compared to NT, genes expressing mRNA in 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7

cell lines were each compared with the those in 66cl4 NT cell line, deducing two lists with calculated statistical values of log_2 fold changes and adjusted p-values (padj) (**Figure 5. A**). Approximately 50 highest ranking genes were first selected from each list which respectively consisted of 786 and 653 genes after being sorted by padj from smallest to largest. Subsequently, genes with log_2 fold changes higher than -1.5 were cut-off.

2.7 ChIP-sequencing

ChIP library preparation (prep) and sequencing service was provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority.

2.7.1 Concentration measurement of ChIP DNA

For quantification of dsDNA in the NRF2-ChIPed replicates, the samples were diluted with fluorescent dyes from Qubit™ dsDNA HS Assay Kit (Cat.: Q32854, Thermo Fisher Scientific) following their recommended protocol [90]. Samples tagged with dyes were subsequently inserted into Qubit® 3.0 Fluorometer (Cat.: Q33216, Thermo Fisher Scientific) for concentration readings.

2.7.2 Library prep

Library prep was performed by GCF following QIAseq Ultralow Input Library Kit (Cat.: 180492, Qiagen) handbook [91]. Mainly six types of libraries were generated which consisted of input libraries, NRF2 ChIP (recombinant anti-NRF2 antibody [EP1808Y]) libraries and

normal rabbit IgG ChIP libraries from each cell line 66cl4 NT and 66cl4 NRF2 KD 3B7. The 66cl4 NT input library was used as background for increased ChIP-seq peak specificity while both 66cl4 NRF2 KD 3B7 cell line and normal rabbit IgG libraries were included as negative controls.

Following library prep, quality control (QC) of the amplified DNA fragments from each library was conducted using an Agilent High sensitivity DNA kit (Cat.: 5067-4626, Agilent Technologies, CA, USA) with Agilent 2100 Bioanalyzer (Cat.: G2939BA, Agilent Technologies) [92].

2.7.3 ChIP-seq

Next generation sequencing (NGS) was conducted by GCF. Libraries were denatured and diluted to a concentration of 2.5 pM using the standard normalization method protocol from NextSeq System Denature and Dilute Libraries Guide (Cat.: 15048776 v09, Illumina, CA, USA) [93]. NGS was thereupon performed [94]. Succeeding NGS, the resulting FASTQ file was analyzed with Multi QC software (v1.9) to ensure the quality of the results and possibly remove false reads [94, 95]. Bioinformatic tools such as FastQC, Fastp and FastQ Screen were utilized in this process which are all supported by Multi QC [95, 96]. FASTQ files contain data which is composed of sequences read from each cluster on a flow cell and is followed by base call quality scores [97]. Base call quality scores are produced during Illumina's sequencing by synthesis process by measuring the fluorescent signal intensities emitted by labeled fluorescent tags on each nucleotide [98]. Based on these readings, a quality (Q) score is generated to predict its reliability and accuracy. High Q scores indicates that the sequences are less likely to be incorrect, whereas low Q scores will imply that they have high error probabilities [99].

2.7.4 Bioinformatic Analyses

2.7.4.1 ChIP-seq analyses with MACS2

The ChIP-seq bioinformatics analyses were performed at the Bioinformatics core facility (BioCore), Norwegian University of Science and Technology (NTNU). BioCore is funded by the Faculty of Medicine at NTNU and Central Norway Regional Health Authority.

To map novel NRF2 binding sites in 66cl4 murine metastatic breast cancer cells, 66cl4 NT samples that were precipitated by anti-NRF2 antibody EP1808Y (NT-NRF2 ChIP) were bioinformatically analyzed. Raw .fastq sequences were aligned to the mouse mm10 reference genome using Bowtie2 [100]. All 24 samples had an alignment rates above 90%, except sample 2 which had 81%. ChIP-Seq enrichment peaks for NRF2 were identified by MACS2 [101] NarrowPeak modus using default setting and FDR cutoff of 0.05. Since very few peaks were identified in the individual replicates, the aligned replicate .bam files in each sample group were pooled, such that MACS2 identified peaks running NRF2 ChIP pooled vs corresponding pooled Input, and NRF2-KD pooled versus corresponding pooled Input. The final peak list was generated after filtering against mm10 blacklisted regions from ENCODE [102].

2.7.4.2 Gene set enrichment analyses

This study used the gene set enrichment analysis (GSEA) method to investigate which biological functions and pathways the genes detected from RNA-seq and ChIP-seq are most associated to. GSEA is a method of interpreting a list of genes to identify the commonly related biological functions or even pathologies. It functions by comparing the genes of interest to accumulated gene data provided from published experimental articles and determining which gene sets the queried genes are significantly enriched in [103]. In this study, this was achieved

by using an open-access online software called Enrichr (https://maayanlab.cloud/Enrichr/) [104, 105].

2.7.4.3 Cancer Cell Line Encyclopedia

The Cancer Cell Line Encyclopedia (CCLE) (https://portals.broadinstitute.org/ccle) was used to search whether the genetic characteristics found in results from RNA-seq experiments exhibit connection with gene expressions in human breast cancer cell lines. CCLE is an online tool created from a project to enable analyzation of genes and its expression levels in human cancer cell lines. The present study mainly identified the mRNA expression levels of the NRF2 encoding gene, *Nfe2l2*, and NRF2-regulated genes in 60 human breast cancer cell lines. Furthermore, mRNA expression levels were compared between *Nfe2l2* and its highly regulated genes in human breast cancer cell lines using a scatter plot [106, 107].

2.7.4.4 cBioPortal

Genetic alteration types and its frequencies found in *Nfe2l2* and NRF2-targeted genes were explored in data sets from large-scale human breast cancer studies and especially within invasive breast cancer studies by utilizing cBioPortal (https://www.cbioportal.org/). cBioPortal is an online open-source tool with accumulated cancer genomic data from sources such as TCGA. The website offers visualization of genetic modifications of queried genes within the selected range of clinical studies. Oncoprint, a horizontal-graph feature of the online tool, was mainly used in this study as it presents the types of genetic alterations by color, and number of patient cohorts by the length of the colored graph [108, 109].

3 Results

3.1 NRF2 is highly expressed in 66cl4 NT cells and effectively suppressed in 66cl4 NRF2 KD cells

To ensure the downregulated expression of NRF2 in 66cl4 NRF2 KD 3B4 and 3B7 cell lines compared to 66cl4 NT, immunoblotting was performed. For NRF2 detection, the NRF2 (D1Z9C) rabbit mAb (Cell signaling Technology) was used as it is known to accurately detect the protein in Western blots compared to other commercial NRF2 antibodies such as EP1808Y (Abcam) and antibodies H-300 and C-20 (Santa Cruz Biotechnology) [110]. NRF2 was expressed in 66cl4 NT cell line with molecular weight of 75 kilodalton (kDa) whilst the band was absent in 66cl4 NRF2 KD 3B7 and very weakly displayed in NRF2 KD 3B4 cell lysates (**Figure 4**, **Supplementary Figure 1**). As loading control, anti-COX IV antibody (Abcam) was selected on account of the generally high expression of the COX IV protein and the dissimilar molecular weight from NRF2, enabling bands to be clearly distinguishable [111]. COX IV was detected with similar intensity in all three cell lines by the 15 kDa band assuring equal loading of proteins in all wells. Taken together, the bands detected with D1Z9C in 66cl4 lysates verify the expression of NRF2 in 66cl4 NT cell line and the efficient KD of the TF in 66cl4 NRF2 KD 3B4 and 3B7 cell lines.

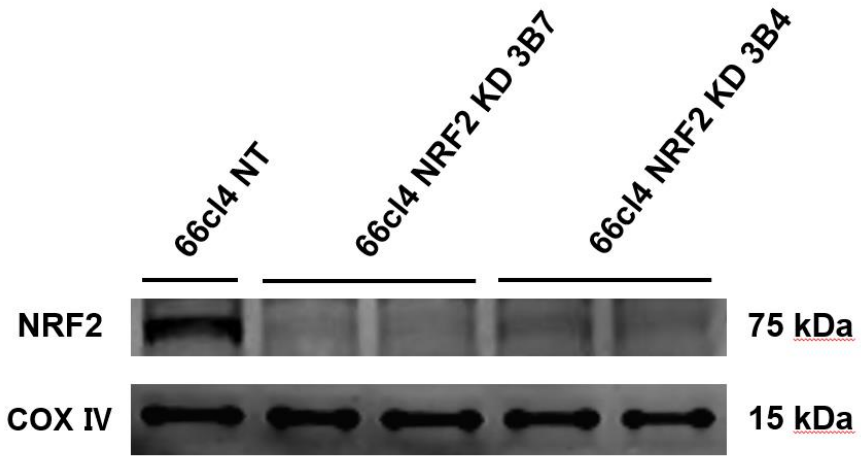


Figure 4. NRF2 is confirmed to be highly expressed in 66cl4 NT cells whereas it is efficiently knocked down in 66cl4 NRF2 KD cell lines. The immunoblot of NRF2 displays an intense band in the 66cl4 NT lane in contrast to the weak bands in 66cl4 NRF2 KD 3B7 and 3B4 cells. COX IV was used as a loading control.

3.2 RNA-seq data analyses of 66cl4 cell lines

3.2.1 Forty genes are commonly downregulated in 66cl4 NRF2 KD cells when compared to 66cl4 NT cells

To predict potential findings of the extracted DNA sequences, RNA-seq data produced by Bjørkøy et al. (unpublished) was analyzed. Consequently, there were 49 and 60 genes with the lowest padj and low log_2 fold change values in lists of repressed genes of 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7. The two lists were contrasted with each other to retrieve commonly downregulated genes (**Figure 5. A**). As a result, there were 40 common genes which are significantly downregulated in NRF2 KD cells compared to NT cells where NRF2 is constitutively expressed (**Supplementary Table 3**).

3.2.2 Commonly downregulated genes in 66cl4 NRF KD cell lines are identified as putative NRF2-targeted genes

Enrichment analysis was conducted on the 40 overlapping genes by Enrichr to explore whether any of these genes are known targets of NRF2 [104, 105]. ChIP-X Enrichment Analysis (ChEA) 2016 gene set, a TF-target gene database integrated from published ChIP and other TF binding site profiling studies, indicated the expression of the queried genes to be regulated by NRF2 (*Nfe2l2*) in mouse embryonic fibroblasts (MEFs). It specifically pointed out genes *Slc48a1*, *Enah*, *Alox5ap*, *Hipk2*, *Zfp652*, *Sim2*, *Tsku* and *Gsta4* (8/40) (**Figure 5. B**) [112]. The Encyclopedia of DNA Elements (ENCODE) and ChEA Consensus TFs from ChIP-X gene set reinforced the result as it indicated that *Slc48a1*, *Enah*, *Alox5ap*, *Hipk2*, *Sim2*, *Tsku* and *Gsta4* (7/40) are targeted by NRF2 (*Nfe2l2*) (**Figure 5. C**) [112-114]. GO Molecular Function 2018 enrichment analysis further showed the participation of *Gstk1*, *Gsta4*, *Alox5ap* (3/40) gene products in glutathione transferase activity (**Figure 5. D**) [115, 116]. Glutathione transferases are phase II detoxification enzymes [117]. As NRF2 is a major TF inducing expression of detoxification enzymes, the findings are in line with rest of the results. Hence, 66cl4 NT cells transcribe NRF2-regulated genes whereas the two NRF2 KD cells do not. Altogether, the results indicate the effective silencing of *Nfe2l2* in cell lines 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7 and confirm that 66cl4 NT cell lines express putative NRF2-regulated genes.

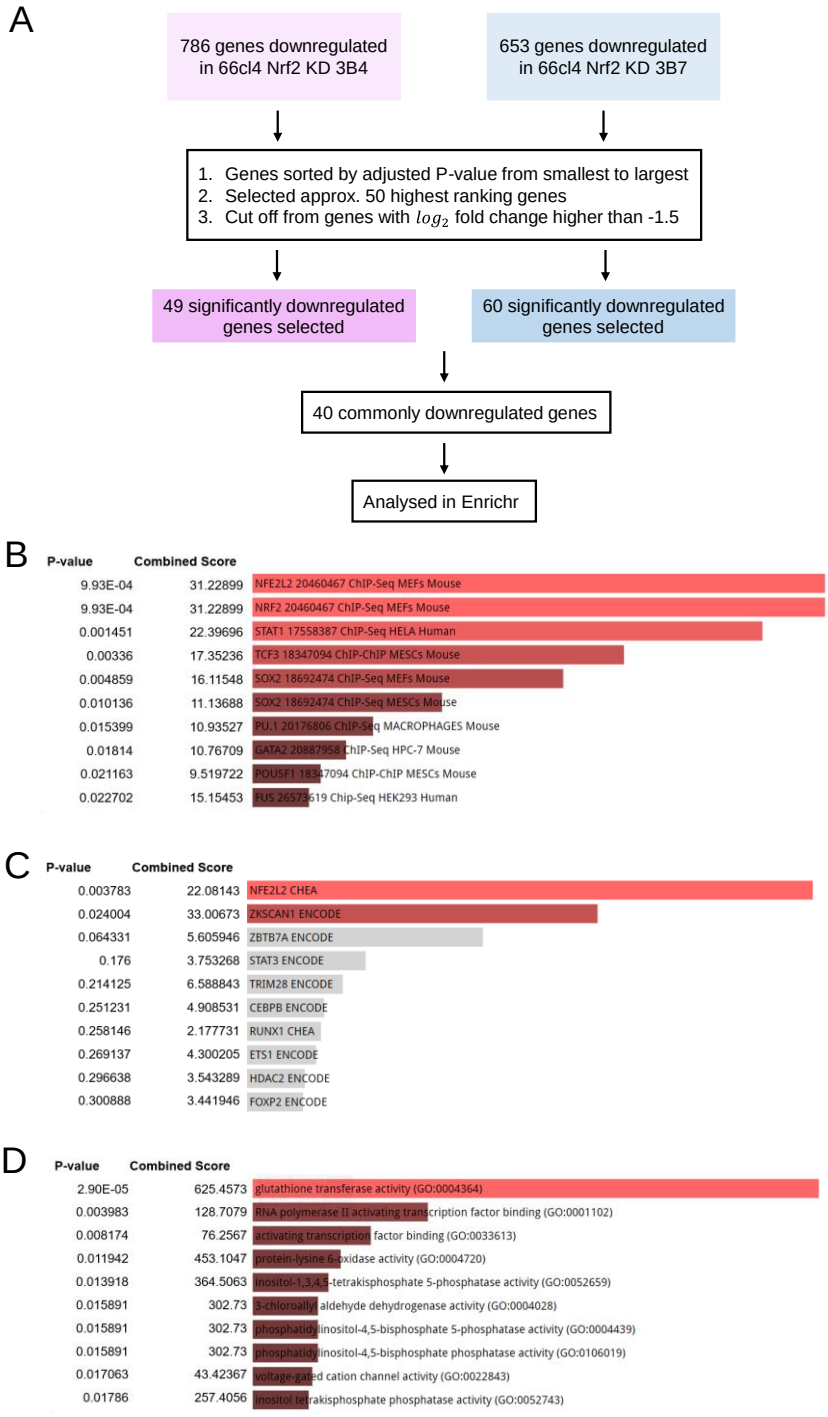


Figure 5. Enrichment analysis of RNA-seq results. Genes differentially expressed in 66cl4 NRF2 KD 3B4 and 3B7 cell lines compared to 66cl4 NT cells. (A) Process of analyzing RNA-seq data of 66cl4 NRF2 KD 3B4 and 3B7 cell lines each with 66cl4 NT cell line as control to determine genes that are effectively downregulated in 66cl4 NRF2 KD cells. Genes with significantly suppressed expressions were sorted according to adjusted P-values and log_2 fold changes. The overlapping genes between the two lists were selected and queried in Enrichr [104, 105]. **(B)** ChIP-X Enrichment Analysis (ChEA) 2016 indicates that the commonly downregulated genes in 66cl4 NRF2 KD cell lines are NRF2-targeted genes [112]. **(C)** ENCODE and ChEA Consensus TFs (TFs) from ChIP-X enrichment analysis reinforces the finding that the suppressed genes are NRF2-regulated genes [112-114]. **(D)** GO

Molecular Function 2018 enrichment analysis shows that some of the commonly downregulated genes are involved in detoxification pathways [115, 116]. Abbreviation: The Encyclopedia of DNA Elements (ENCODE), ChIP-X Enrichment Analysis (ChEA) and Gene Ontology (GO).

3.2.3 mRNA expression correlation of *Nfe2l2* and putative NRF2-target genes in breast cancer cell lines

To examine the expressional correlation of NRF2 and its target genes in breast cancer, mRNA expression of the seven genes retrieved from Enrichr analysis was analyzed using mRNA expression data of 60 breast cancer cell lines through Cancer Cell Line Encyclopedia (CCLE) [106, 107]. mRNA expression of *ENAH*, *HIPK2*, *TSKU*, *GSTA4* (4/7) exhibited a close proportional relationship with expression of *NFE2L2*. This further confirms the highly expressed NRF2 activity in the metastatic breast cancer 66cl4 NT cell line in contrast to the two NRF2 depleted cell lines (**Supplementary Figure 2**).

3.2.4 Mutations in NRF2 target genes found in human breast cancer tissues

To determine the link between NRF2 and breast cancer development, the known NRF2-targeted genes found in RNA-seq were investigated to check whether their genetic alterations were identified in human breast cancer biopsies. cBioPortal was used to identify the genetic modifications of *SLC48A1*, *ENAH, ALOX5AP, HIPK2*, *SIM2*, *TSKU* and *GSTA4* in all subtypes of human breast cancer samples retrieved from 16 studies consisting of 6408 breast cancer patients [108, 109]. There were 17% (1119/6408) of patients with alterations in the abovementioned genes which were mostly amplifications with especially high quantities found in *ENAH* and *TSKU* (**Figure 6. A**). The combined studies were then narrowed down to 11 invasive breast cancer studies to consider the potential role of NRF2 in metastaticity. According to cBioPortal, there were 24% (1113/4548) of invasive breast cancer patients with

alterations in the NRF2-targeted genes (**Figure 6. B**). The percentage of patients with genetic modifications was higher in the invasive group than in the total breast cancer group. The studies not only indicate the aberrated role of NRF2 in general breast cancer, but also in aggressive breast cancer development.

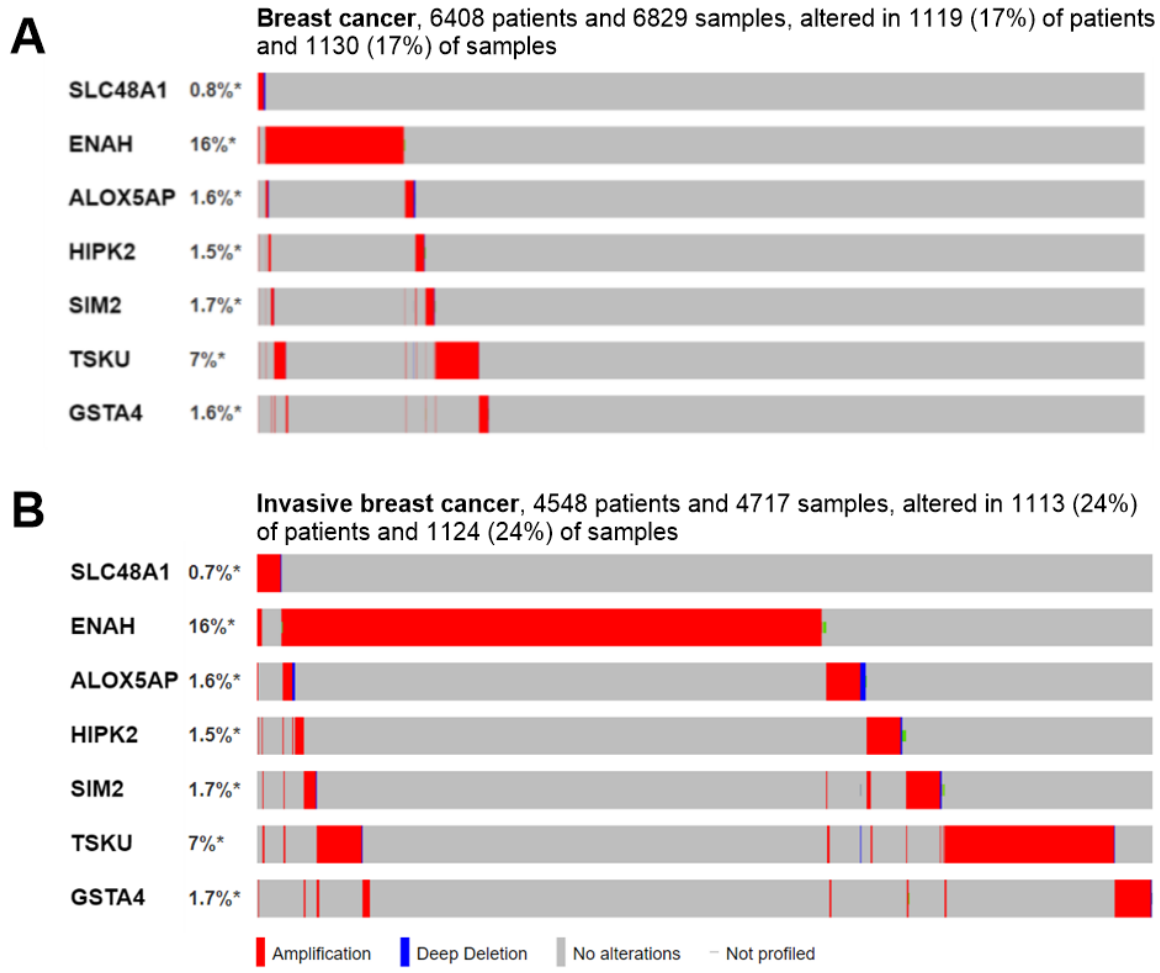


Figure 6. NRF2-regulated genes of 66cl4 cells found in RNA-seq data are highly mutated in clinical breast cancer samples according to cBioPortal database. (A) Mutation status of known NRF2-targeted genes in breast cancer biopsies from 16 clinical breast cancer studies are shown. **(B)** Mutation rates of the NRF2-targeted genes were even higher in invasive breast cancer samples from 11 studies. Most were characterized with gene amplification mutations (red bars) [108, 109].

3.3 ChIP protocol optimization

3.3.1 Twelve sonication cycles shears chromatin in 200-500 base pairs fragments which is a suitable length to be used in both ChIP-qPCR and ChIP-seq

To find the appropriate sonicating cycle number in generating suitably sized DNA fragments, input and pellet samples from each cell line were sheared in various numbers of sonication cycles and examined with agarose gel electrophoresis (**Figure 3. D, E**). The chromatins were aimed to be sheared approximately in 200-500 base pairs (bp). This was for the fragment lengths to be optimized for both qPCR and ChIP-seq. Lysates from cell lines 66cl4 NT, 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7 were sonicated in 5, 10, 12 and 14 cycles (30 sec ON, 30 sec OFF) at 4 °C with Bioruptor Pico sonication device (Diagenode). Sonicated samples were subsequently reverse-crosslinked, electrophoresed and analyzed under UV illumination to identify the adequate cycle number which shears chromatin within the targeted length. According to the UV scanned gels, 12 cycles were the most optimal sonication repeat for all cell lines as the 12 cycle gel bands showed darker shades between 200 and 500 bp compared to other cycle numbers. Thus, all the subsequent biological samples were sonicated in 12 cycles (**Figure 7**).

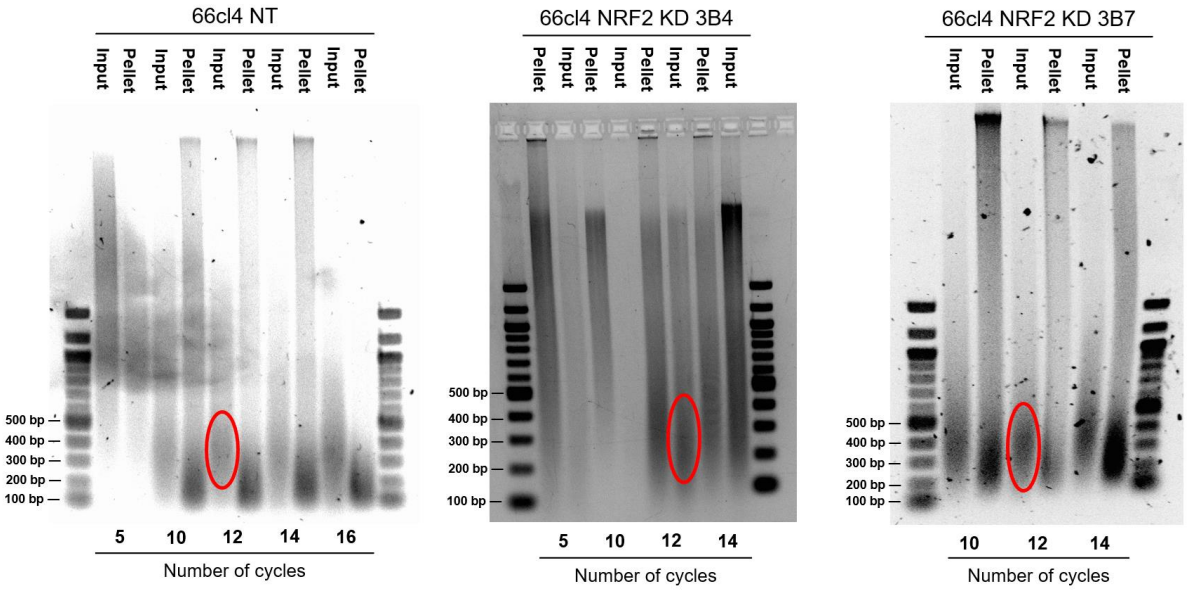


Figure 7. Sonication cycle number optimized to 12 cycles in all 66cl4 cell lines after visualization with agarose gel electrophoresis. Chromatin lysates from cell lines 66cl4 NT, 66cl4 NRF2 KD 3B4 and 66cl4 NRF2 KD 3B7 were each fragmented in different number of cycles (30 sec ON, 30 sec OFF) by Bioruptor Pico (Diagenode) to discover the appropriate cycle number for shearing the DNA into sizes from 200 to 500 base pairs (bp). The fragmented DNA were presented by agarose gel electrophoresis for size identification. The input bands sonicated by 12 cycles were located within the targeted size in all three cell lines. Hence, 12 cycles was selected as the optimal sonication cycle number.

3.3.2 Recombinant anti-NRF2 antibody (EP1808Y) selected as the optimal NRF2-targeting antibody

Selection of an optimal antibody for ChIP-seq is crucial as its specificity will be a significant factor in highly enriching genes of interest (**Figure 3. F-a**). A sensitive antibody will precipitate large quantities of chromatin fragments crosslinked with the targeted protein, while antibodies with low specificity will retrieve background noises hindering the specific detection [118]. In this study, two NRF2 antibodies were used and compared for optimalization to produce samples with highly enriched NRF2-DNA complexes. Recombinant anti-NRF2 antibody (EP1808Y) from Abcam and NRF2 (D1Z9C) rabbit mAb from Cell Signaling Technology were carefully selected as several studies which had used one of the two, validated their effectiveness against NRF2 [22, 87, 110, 119].

To compare the ability of EP1808Y and D1Z9C, both antibodies were separately applied to precleared chromatin samples. Following the procedure, samples were evaluated by qPCR with primers of five known NRF2-targeted genes including *Nqo1* and *Txnrd1*. Primer sequences are shown in **Supplementary Table 1** and qPCR results for all five sequences can be found in **Supplementary Figure 3** [87]. Two sets of biologically replicated ChIP samples were prepared for qPCR examination. From the two qPCR results, EP1808Y yielded higher quantities of NRF2-regulated genes in most samples than D1Z9C. The percent input value of C_T from EP1808Y results were in average 10.34 times and 1.30 times higher than D1Z9C in the first and second set of ChIP-qPCR, respectively (**Figure 8**). The results of negative control normal rabbit IgG from Cell Signaling Technology maintained a low status in all three cell lines. This was foreseen as the antibody is non-specific to known antigens [83]. Thus, EP1808Y was selected to be used for the remaining ChIP performance as it exhibited higher specificity than D1Z9C.

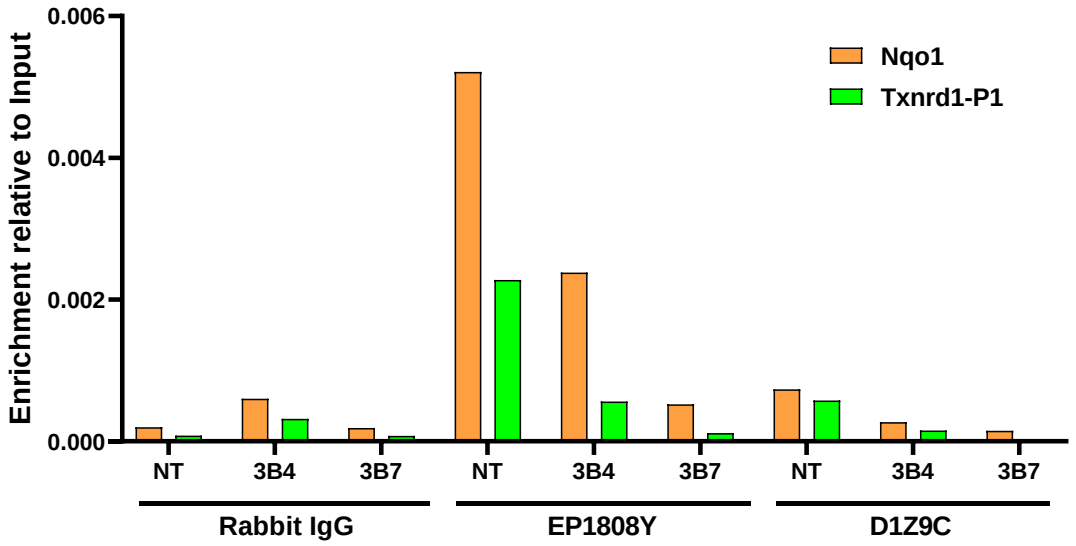


Figure 8. Recombinant anti-NRF2 antibody (EP1808Y) (Abcam) exhibited higher specificity than NRF2 (D1Z9C) rabbit mAb (Cell Signaling Technology) in ChIP samples. EP1808Y showed distinctly stronger enrichment signals than D1Z9C in 66cl4 NT cells when examined with primers of NRF2-regulated genes which were referenced from an NRF2 ChIP-seq study by Malhotra et al. [87]. Negative control normal rabbit IgG (Cell Signaling Technology), on the other hand, induced low signals as anticipated. Amplified DNA levels were calculated by percent input method using C_T values of input samples for normalization [89]. The graph presents average values from the first and second set of ChIP-qPCR. Abbreviation: Txnrd1-P1 (Txnrd1), 66cl4 NT (NT), 66cl4 NRF2 KD 3B4 (3B4) and 66cl4 NRF2 KD 3B7 (3B7).

3.3.3 Primers of genes *Nqo1* and *Txnrd1* selected as optimal primers for validating the quality of ChIP samples

In advance of conducting high-throughput DNA sequencing, it is important to validate the quality of the ChIPed DNA. This is commonly done by qPCR assays [67]. To efficiently examine its quality, primers that exhibit high qPCR amplification signals were required. Thus, primers that are optimal for assessing the enrichment of NRF2-targeted genes had to be determined ahead of ChIP-DNA validation. For this experiment, a positive control primer and several primers of known NRF2-binding sites were prepared, examined, and compared by

qPCR. The primers of putative NRF2-regulated genes were referenced from a study by Malhotra et al. (**Supplementary Table 1**) which contained a list of primers they had organized for PCR assays based on the genes they had found from their Nrf2 ChIP-Seq peak results [87]. From the list, five primers were specifically selected, as several articles have reported them to be NRF2-downstream targets that express antioxidant or detoxification enzymes [63, 120, 121]. Among them were primers for genes *Nqo1* and *Txnrd1*. All primers were used for target gene amplification in input and IP samples. qPCR was performed twice for two biologically replicated ChIP samples and percent input values were subsequently calculated [89].

In the two 66cl4 NT samples enriched by recombinant anti-NRF2 antibody (EP1808Y) (Abcam), *Nqo1* level was significantly elevated in both first and second set of qPCR experiments (**Figure 8**). Although *Txnrd1* levels were lower than *Nqo1*, *Txnrd1* also resulted in considerable quantities. The rest of the three genes, however, did not show significant enrichment signals compared to *Nqo1* and *Txnrd1* (**Supplementary Figure 3**).

For samples amplified with positive control SimpleChIP mouse RPL30 intron 2 primers (Cell Signaling Technology), positive control Histone H3 (D2B12) rabbit mAb (Cell Signaling Technology) IP samples showed significant percent input values ranging from 0.165 to 0.298, whereas values from remaining IP samples ranged from 0 to 0.002 (**Figure 9**). This was as expected since Histone H3 mAb would pull down considerate amounts of *RPL30* genes, whereas antibodies of other IP samples would not. *RPL30* is a housekeeping gene that is actively transcribed in all cell types and Histone H3 is a protein bound to most DNA sequences. Thus, the positive control mAb would enrich *RPL30* genes by targeting Histone H3 [85, 88]. For negative controls, concentration levels of the genes both in 66cl4 NRF2 KD samples and in normal rabbit IgG (Cell Signaling Technology) IP samples were distinctively low as expected.

Taken together, ChIP-qPCR assays showed primers Nqo1 and Txnrd1 to be the most optimal in evaluating the quality of NRF2-ChIP samples and further validated the implemented ChIP method of this study by using control samples.

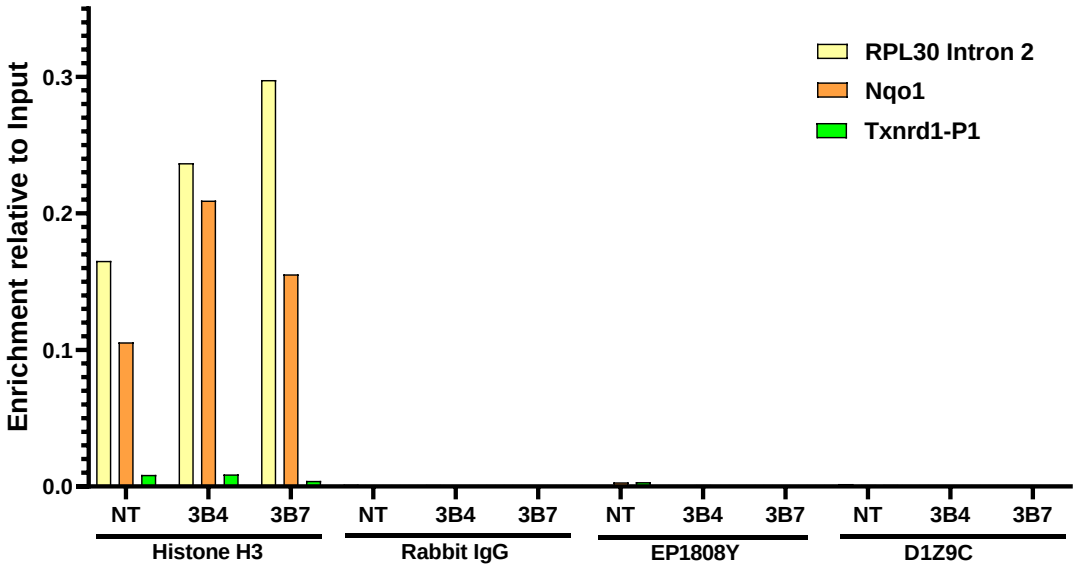


Figure 9. ChIP-qPCR results indicate the effectiveness of the ChIP method practiced in this study. ChIP DNA immunoprecipitated with Histone H3 (D2B12) rabbit mAb (Cell Signaling Technology) yields high levels of amplified *RPL30* genes in ChIP-qPCR. Meanwhile, *RPL30* gene signals can scarcely be detected in other negative control (Rabbit IgG) and target (EP1808Y, D1Z9C) samples. Since *RPL30* is a house-keeping gene and Histone H3 are one of the core proteins forming all chromatin structures in eukaryotic cells [85], it was expected for the *RPL30* gene to generate high signals in Histone H3-ChIP-qPCR and extremely low signals in other antibody-ChIP-qPCR samples. Enriched DNA levels were measured using percent input method and normalized by C_T values of input samples [89]. Abbreviation: Txnrd1-P1 (Txnrd1), 66cl4 NT (NT), 66cl4 NRF2 KD 3B4 (3B4), 66cl4 NRF2 KD 3B7 (3B7), Normal Rabbit IgG from Cell Signaling Technology (Rabbit IgG), anti-NRF2 antibody EP1808Y from Abcam (EP1808Y) and NRF2 D1Z9C rabbit mAb from Cell Signaling Technology (D1Z9C).

3.4 ChIP-sequencing

3.4.1 Five ChIP biological replicates prepared for ChIP-seq

The aim of this project was to map genome-wide NRF2-binding sites in 66cl4 cells with maximal signal-to-noise ratio. However, variability and noise in samples can lead to weak results [70]. An approach to minimize the impediment is preparing biological replicates to improve the accuracy and reliability [122]. This method can strip false findings, whereas genuine findings are emphasized as they will exhibit consistency among the replicates [70]. To produce significant results holding generalizability in 66cl4 metastatic breast cancer cells, biological ChIP replicates were organized to be used for library prep and NGS.

After optimizing the ChIP assay, a total of six sets of biological replicates were first prepared by repetitively conducting the ChIP procedure with 66cl4 cell lines that were individually harvested in different time points from three separate batches of cells. To assess the quality of the replicates, qPCR was executed in all six ChIP DNA replicates with primers Nqo1 and Txnrd1. Most replicates were immuno-enriched with *Nqo1* and *Txnrd1* in 66cl4 NT samples while the gene levels were low in 66cl4 NRF2 KD 3B4 and 3B7 samples (**Figure 10**). In general, 66cl4 NRF2 KD 3B7 cells presented more effective knock down of NRF2 than 66cl4 NRF2 KD 3B4. The 66cl4 NRF2 KD 3B4 cell lines was thereupon excluded for qPCR examination in the latter two replicates (**Figure 10. B**) and further omitted from library prep followed by sequencing. Moreover, most negative control normal rabbit IgG ChIP samples exhibited percent input values close to zero. Overall, 66cl4 NT cells displayed higher enrichment levels of NRF2-ChIP DNA than 66cl4 NRF2 KD cells in all six replicates and thereby were verified to be adequate in further proceeding concentration measurements for library prep.

Concentration of each replicate was measured with Qubit™ dsDNA HS Assay Kit and Qubit® 3.0 Fluorometer to select the five most dsDNA-enriched samples [90]. Concentrations are shown in **Supplementary Table 4** in the appendices. Amongst the six replicates, all EP1808Y-ChIP and Rabbit IgG-ChIP samples within the first ChIP replicate could not be detected due to low readings. The concentration detection range of the dsDNA assay kit is between 10 pg/µl and 100 ng/µl which conveys that the samples with low readings had concentrations below 10 pg/µl [90]. Moreover a minimum of 1 - 10 ng of ChIP DNA is advised for library construction [123]. Consequently, it was decided to use the rest of the five replicates as the first replicate was not suitable to be used for library prep.

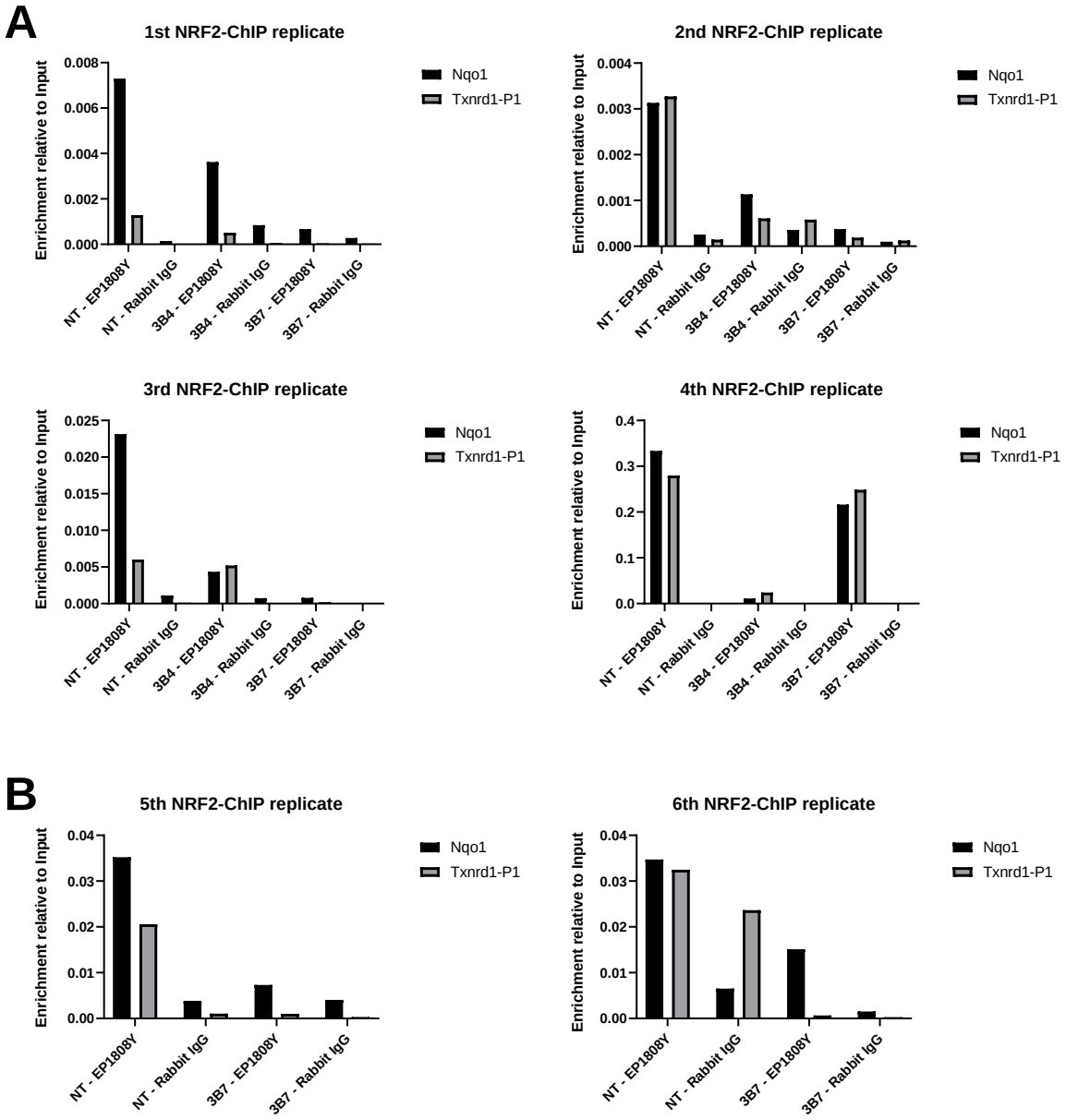


Figure 10. All 66cl4 NT ChIP replicates confirmed to be enriched in gene fragments targeted by NRF2. A total of six 66cl4 biological replicates were prepared and validated by ChIP-qPCR to be enriched with NRF2-ChIP DNA in 66cl4 NT (NT) samples, whilst being selectively depleted in 66cl4 NRF2 KD 3B4 (3B4) and 66cl4 NRF KD 3B7 (3B7) samples. All anti-NRF2 antibody EP1808Y-treated NT samples showed significant enrichment of the two NRF2-regulated genes which were *Nqo1* and *Txnrd1*. Conversely, all samples that were precipitated with the negative control normal rabbit IgG, maintained low concentration of the two NRF2-targeted genes. **(A)** In addition, 3B7 cells generally presented more effective knock down of NRF2 than 3B4. **(B)** 3B4 was thereupon excluded for qPCR examination in the latter two replicates.

3.4.2 Quality control verifies adequacy of ChIP-library for NGS

Following biological replicate selection, ChIP libraries of the five replicates were prepared. Once the libraries were denatured and diluted to 2.5 pM using the Standard Normalization method from Illumina NextSeq System guideline [94], a QC was performed. This was to assess adapter ligation, confirm amplification, and identify concentration and size distribution of DNA fragments. Using Agilent High sensitivity DNA kit, a DNA chip loaded with gel-dye mixes, markers, and samples were run and examined with 2100 Bioanalyzer instrument (Agilent Technologies) [92]. Most wells produced highly concentrated libraries and the lengths of ChIP-DNA were within the targeted range (**Figure 11**). The average concentration for each detected peak was 4.67 nM. The recommended minimum library concentration prior to sequencing is 0.5 nM which means the constructed library contains enough quantities of ChIP-DNA for sequencing [94]. Furthermore, the detected average insert length was 448.17 bp which implies the inserted DNA length to be 328.17 bp given that the sizes of Qiagen adapters are 120 bp [91]. As this study had sheared DNA in a length of 200 – 500 bp, the bioanalyzer result identified adapter ligation and re-confirmed the fragmented DNA length. Taken altogether, library evaluation confirmed PCR-amplification of adapter-ligated samples and further verified the library samples to be adequate in both concentration and size distribution for ChIP-seq.

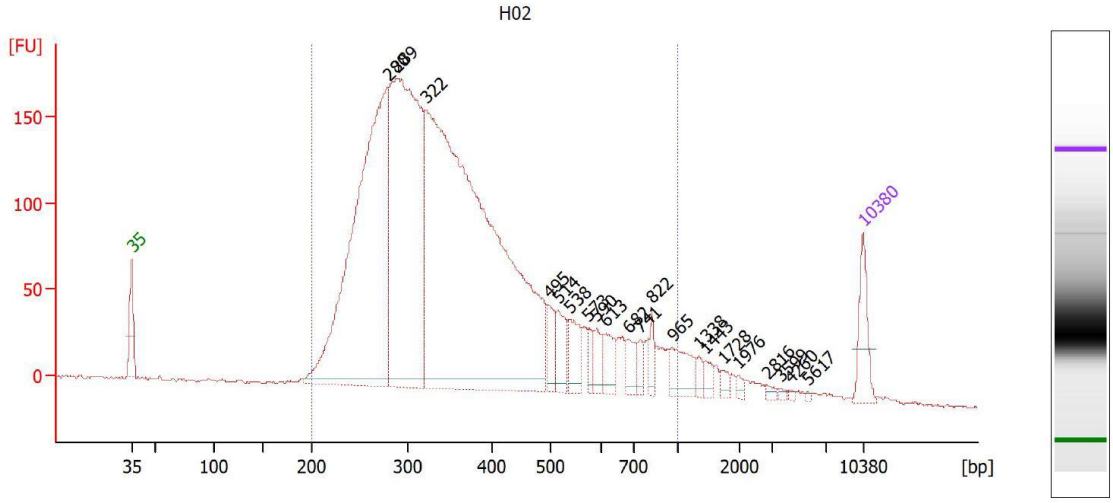


Figure 11. An electropherogram of EP1808Y-ChIPed 66cl4 NRF2 KD 3B7 sample exhibits features of a successful ChIP-library prep. The displayed electropherogram is a representative example from the 5th biological replicate among the graphs from other libraries. The electropherogram obtained from 2100 Bioanalyzer (Agilent Biotechnologies) shows a clear peak in the 300-base pair (bp) region with a concentration of 1310.01 pg/µl (not shown in figure) which suggests that it meets the requirements for subsequent Illumina sequencing [94, 124]. The graph also clearly displays both lower (green) and upper (purple) marker peaks. While sample peak in distinctly situated in between the markers, rest of the regions appear to have a flat baseline which altogether indicates a high-quality library [92]. Abbreviation: Recombinant anti-NRF2 antibody EP1808Y from Abcam (EP1808Y).

3.4.3 Multi QC validates quality of ChIP-Seq data

For identification of the NRF2-occupied genomic regions in 66cl4 cells, the quality-verified libraries underwent massive parallel sequencing by the NTNU's Genomics Core Facility (GCF) which produced a FASTQ file of the enriched sequences [94]. The FASTQ file was thereafter analyzed with Fastp to ensure the quality of the results and possibly remove false reads. Fastp is a module that parses FASTQ files through QC and data-filtering [96]. This can be run by Multi QC software (v1.9) [94, 95]. According to Fastp, the vast majority of reads from each sample were above the quality filter whereas low quality reads and short-fragment reads were limited to a small percentage. (**Figure 12. A**). Moreover, the ChIP-library had amplified an adequate amount of reads approximately ranging from 6 million (mil) to 20 mil. As the aim

was generating 10 to 20 mil reads with high quality, the results implied a successful ChIP library prep and sequencing.

Furthermore, the ChIP-sequences were mapped to a set of reference genomes to determine its quality by validating its origin. The method is recommended to find possible sources of contamination. The analysis was performed with a bioinformatic tool called FastQ Screen which is also integrated in MultiQC [95, 125]. The tool first quantifies the user's input reads and compares the data against a large sequence dataset extracted from a variety of reference genomes [125]. The results generated by FastQ Screen showed that ChIP-sequences largely matched with databases from mice and was followed by multiple genome databases (**Figure 12. B**). As the ChIP-DNA samples were immunoprecipitated from 66cl4 murine cells, this not only verified the origin of the sample, but also its quality.

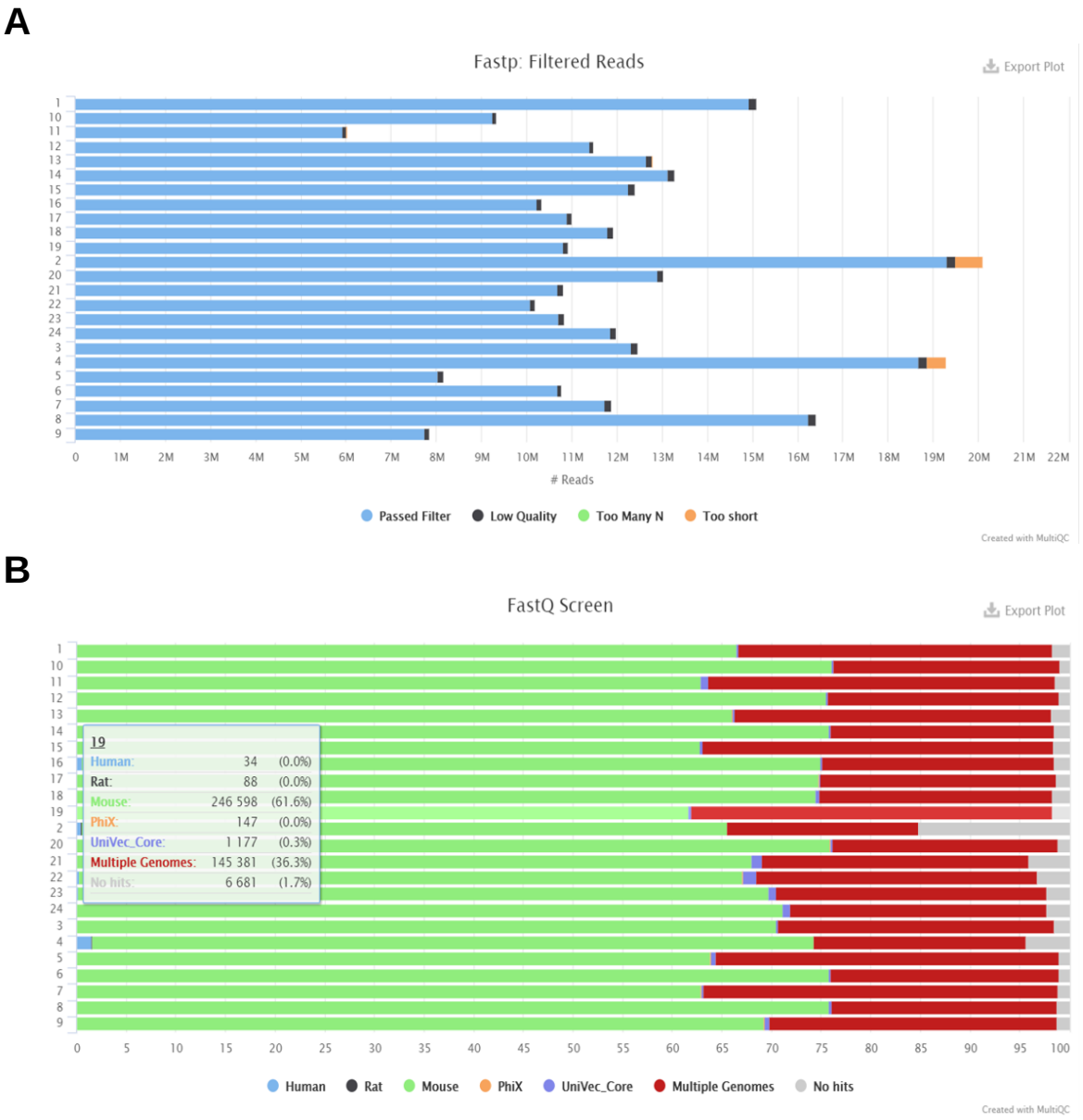


Figure 12. Quality of the 66cl4 ChIP-seq data was verified by bioinformatic tools implemented in Multi QC software (v1.9). (A) The library of ChIP-sequences was validated by Fastp to have high quality which is indicated with blue-colored bars (passed filter) in the graph. The quality was determined by screening quality scores of each read. The library also generated a number of reads which reaches the target range of 10 to 20 million [96]. **(B)** The graph created from FastQ Screen additionally confirms high quality of the samples [125]. As expected, the bars show that the ChIP samples purified from murine 66cl4 cells mostly match with reference genomes from mice (green) and multiple genome (red) datasets [95].

3.4.4 Global identification of NRF2 binding sites in 66cl4 murine metastatic breast cancer cells

3.4.4.1 Bioinformatic ChIP-seq analyses finds 94 significant peaks enriched by anti-NRF2 antibody in 66cl4 NT and one significant peak in 66cl4 NRF2 KD 3B7 cells

To obtain a list of significant NRF2-modulated genomic regions in 66cl4 cells, bioinformatic analyses were conducted by the Bioinformatics Core facility (BioCore) of NTNU. The results for each sample group were pooled for analysis which generated a total of 118 peaks. Once noise signals were filtered, 94 significant ChIP-seq peaks were identified in 66cl4 NT samples that were precipitated by anti-NRF2 antibody EP1808Y (NT-NRF2 ChIP) (**Figure 13**, **Supplementary Table 5**). However, only one significant peak was found in 66cl4 NRF2 KD 3B7 samples (**Supplementary Figure 5**). The results confirmed our expectations. As 66cl4 NT cells maintain constitutive expression of NRF2, the cells ChIPed with anti-NRF2 antibodies were anticipated to extract NRF2-regulated DNA fragments. Negative controls, on the other hand, are either depleted with NRF2 or its DNA could not be precipitated due to non-specific binding of normal rabbit IgG. Thus, it was expected to detect only a trivial number of peaks in those libraries. The bioinformatic results not only confirmed the high efficiency of NRF2 KD in the 66cl4 NRF2 KD 3B7 cell lines, but also validated the applied ChIP-seq methodology and its success.

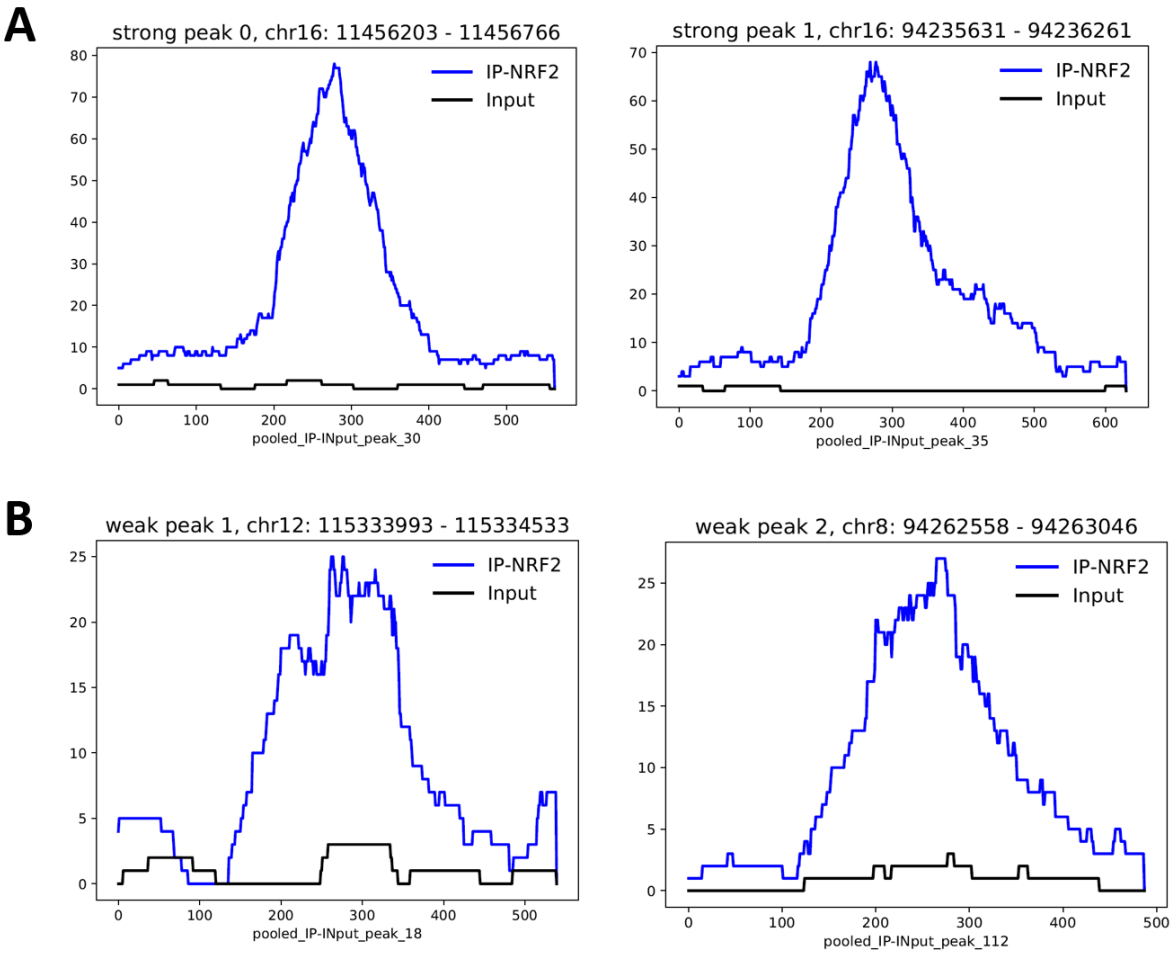


Figure 13. A total of 94 significant ChIP-seq peaks were identified in anti-NRF2 antibody-enriched 66cl4 NT libraries (NT-NRF2 ChIP). (A) The two most significant ChIP-seq peaks among the 94 are displayed. Both peaks present a peak height of approximately 70 to 80 (blue line) while the background 66cl4 NT input data (NT-input) maintain flat (black line) indicating the sequenced peak region to be highly enriched in NT-NRF2 ChIP samples but not in the background sample. **(B)** ChIP-seq peaks are featured with an approximate height of 25 and thus is considered as the least significant from the 94 peaks. However, NT-NRF2 ChIP samples are clearly distinguished from NT-input samples in terms of enrichment levels. All peaks presented in this figure were obtained from the Bioinformatics Core Facility of NTNU.

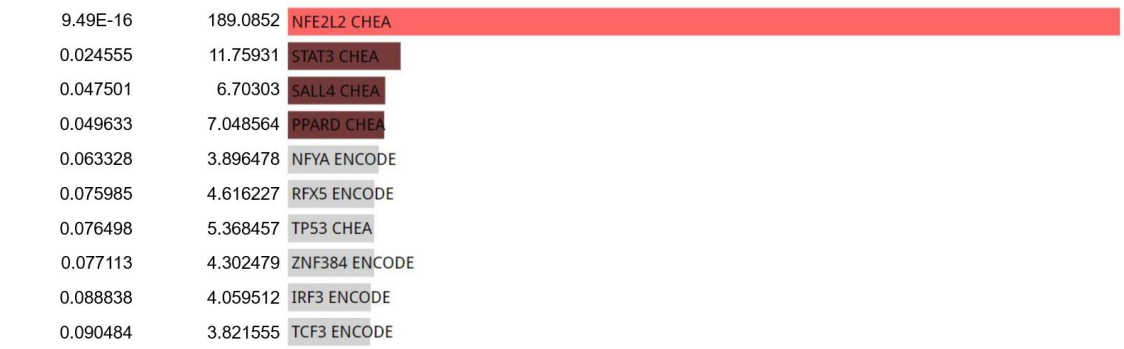
3.4.4.2 ChIP-seq peaks are validated to be regions regulated by NRF2

To verify that the ChIP-seq peak regions are NRF2-regulated sites, ChIP-seq peaks were examined on the UCSC genome browser and observed whether the peaks were located in vicinity of known NRF2-regulated genes (**Supplementary Figure 4**) [126]. The list of 94 significant peaks were sorted by significancy of $-\log 10\ (Q\ value)$ as recommended by BioCore. Each peak region was subsequently identified through the browser by investigating the type of gene or genomic site it occupies. Genes that are both neighboring and located downstream from the significant peaks were perceived as NRF2-regulated genes. However, not all peaks were enriched contiguously to annotated genes. For those peaks, genes located within a window length of 1,000,000 bp (1 Mbp) were additionally discerned. The bp length was determined to encompass the fact that enhancers are long-distance regulatory elements that can be situated up to 1 Mbp upstream of a gene [126]. This brought an enumeration of 249 genes situated in proximity of the peaks (**Supplementary Table 5**). The 249 genes were thereupon queried on Enrichr to verify if they have been identified as NRF2-regulated genes in previous literatures [104, 105, 127].

As a result, a total of 40 genes (40/249), neighboring 35 peaks (35/94), were confirmed to be regulated by NRF2 according to the ENCODE and ChEA Consensus TFs from ChIP-X gene dataset which was accessed through Enrichr (**Figure 14**). ENCODE and ChEA Consensus TFs from ChIP-X is a gene dataset presenting consensus target genes for various types of TF. In their NFE2L2 (NRF2) data, 40 NRF2-targeted genes were found to overlap with our list of significant ChIP-seq genes. This indicates that the 40 genes (40/249) are known-NRF2-regulated genes which is verified by both ENCODE and ChEA (**Figure 14. B**). The 40 genes were marked with a significant P-value of 9.49E-16 and combined score of 189.09 (**Figure 14. A**) [112-114].

On the contrary, genes in proximity of the remaining 59 peaks (59/94) were not established as NRF2-targeted genes in the public database [104, 105, 112-114]. The 59 peaks can be inferred that these are NRF2-interacting AREs that is yet to be identified. Furthermore, 46 peaks were characterized with distal-enhancer-like signatures and 4 peaks with proximal enhancer-like signatures according to the ENCODE Candidate Cis-Regulator Elements dataset [128]. With more than half of the peaks (50/94) being featured in enhancer signatures, this reinforces our expectancy that these peak regions are precipitated from a TF. Altogether, the results corroborate the list of peak regions to be NRF2-regulated sequences and suggest the 59 regions within the list to be potential novel NRF2 binding sites.

A



B

SLC48A1	RPS15A	GSTM3	NR2F2	PIGK
SLC44A3	BMPER	GSTM2	TERF2	CD9
SYNPO2	FTH1	ABCC2	SRP9	BLVRB
SLC40A1	TMEM108	GSTM1	ENAH	CRELD1
MGST2	PAK6	CAMK1D	SLC22A23	UTRN
SEMA3E	HMOX1	FNBP1	BFSP2	SEC24D
AK5	EMC7	MAGI3	RASA4	TKT
SNX29	POLR2J	EPHX1	TBCE	SIM2

Figure 14. Forty genes located by 35 significant peaks were identified as NRF2-regulated genes. **(A)** Genes proximate to the 94 significant ChIP-seq peaks were searched on ENCODE and ChEA Consensus TFs from ChIP-X gene dataset through Enrichr which confirmed the genes to be targeted by NRF2 (*NFE2L2*). **(B)** A total of 40 genes overlapped with the gene dataset [112-114]. Abbreviation: The Encyclopedia of DNA Elements (ENCODE), ChIP-X Enrichment Analysis (CHEA).

3.4.4.3 ChIP peak list of known-NRF2-regulated sites concur with list of putative NRF2 targets

To further establish the findings that the genes in proximate locations of the 35 ChIP-seq peaks are known-NRF2-targets, they were contrasted with other putative NRF2-modulated genes. A list of ten NRF2-dependent transcriptional targets were selected from an NRF2 ChIP-seq study performed by Malhotra et al [87]. Among the ten genes, six genes were shown to be overlapping with our dataset. The genes were *Hmox1*, *Gstm1*, *Gstm3*, *Nqo1*, *Txnrd1* and *Ephx1*. As an example, in our data, two statistically significant peaks were located upstream within 10 kilobase pairs (kbp) from *Hmox1*, a renowned NRF2-responsive gene known for expressing anti-oxidases (**Figure 15. A**) [129]. Peak heights of the two enriched regions were 57 and 45. Peak heights reflect the number of ChIP DNA fragment tags overlapping in the specific site [130]. A peak with a height of 47 was also detected within the intronic region of *Gstm1* and 45 kbp upstream of *Gstm3* (**Figure 15. B**). Both genes express glutathione S-transferases which are major phase II detoxification enzymes that detoxify carcinogens and metabolize xenobiotics [117].

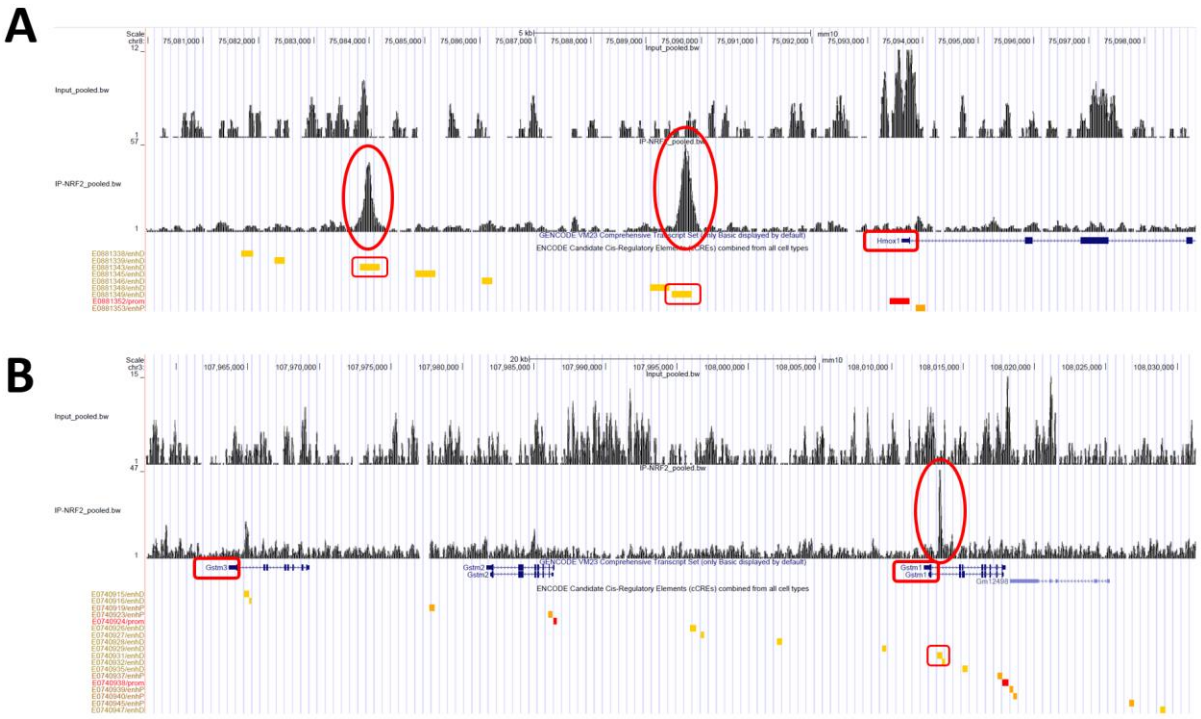


Figure 15. Strong significant ChIP-seq peaks were observed in enhancer regions close to putative NRF2-regulated genes. (A) Two enriched peaks with an average height of 51 were detected in proximity to gene *Hmox1*, a well-known NRF2-regualted gene [129]. Both peaks had distal enhancer-like signatures as indicated with a yellow bar underneath the peaks. In contrast, the input peaks, presented in top row, only had a highest peak of 12. **(B)** A peak with a height of 47 was also detected with a distal enhancer-like signature located within the intronic area of *Gstm1* and upstream of *Gstm3*. Both genes are known to be NRF2 modulated. Its input peaks, on the other hand, exhibited a peak height of 15. The peak datasets were displayed with UCSC genome browser [126].

In addition to the list of 94 significant ChIP-seq peaks, peaks that were not incorporated in the list but nevertheless presented prominent peak heights were also found. Two of those peaks were located by genes *Nqo1* and *Txnrd1* (**Figure 16. A, B**). The two genes are distinguished NRF2 targets which both translate antioxidant enzymes preventing production of ROSs [131, 132]. Despite being excluded from the list, each peak was distinctly structured compared to its background input data. The peaks by *Nqo1* and *Txnrd1* moreover were characterized with proximal enhancer-like and promoter-like signatures, respectively. These finding corresponded to previous qPCR results where elevated enrichment of genes *Nqo1* and *Txnrd1* were limited to NT-NRF2 ChIP samples (**Figure 16. C**). Taken together, overlapping findings between the

analyzed NRF2 ChIP peaks and the putative NRF2 target genes clearly indicates that the peak regions are NRF2-binding sequences.

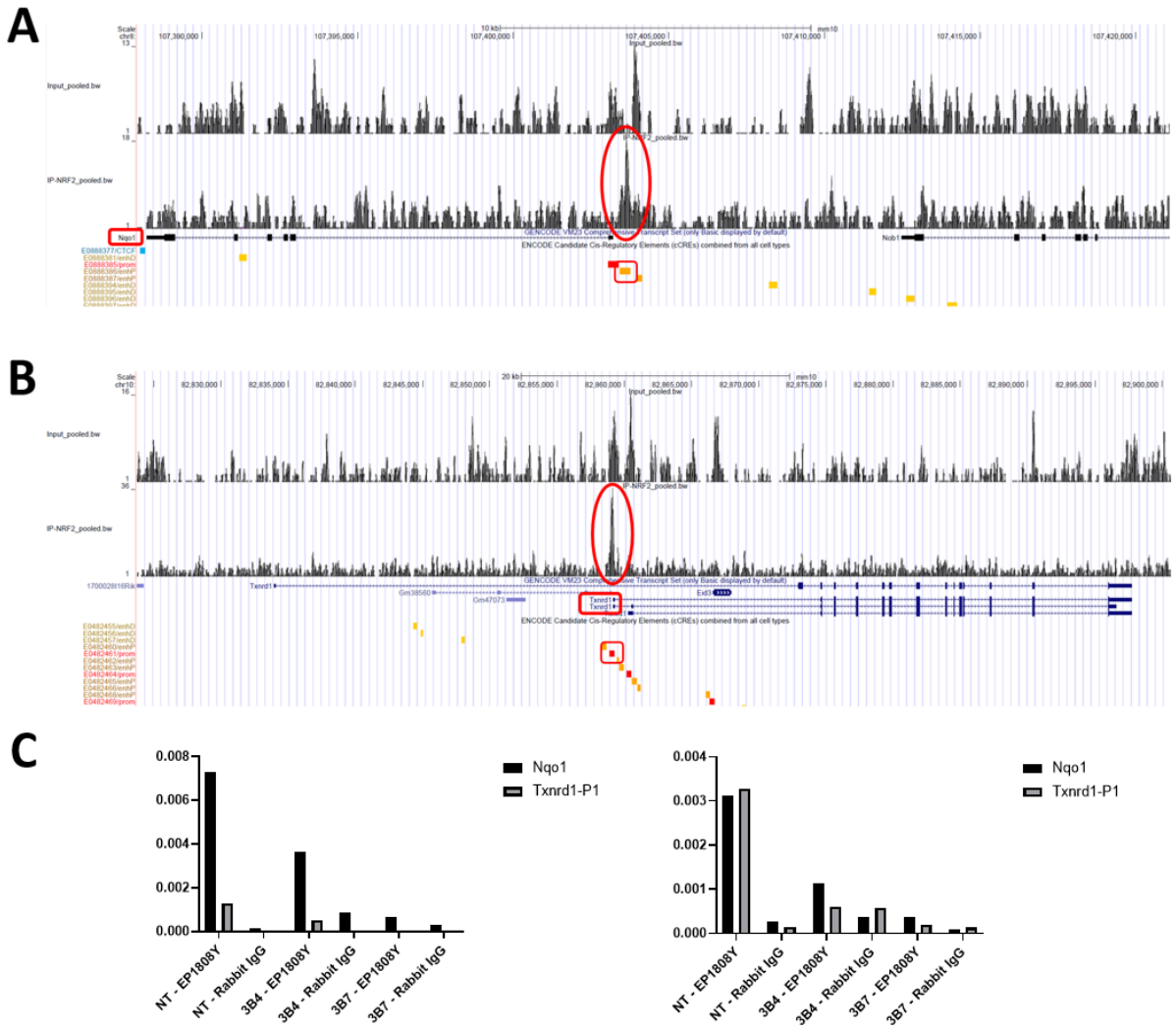


Figure 16. Some elevated ChIP peaks were not included in the list of 94 significant peaks but were situated adjacent to renowned NRF2-regulated genes. (A) A distinct peak with a proximal enhancer-like signature (orange bar) was found neighboring the transcription start site of *Nqo1*. **(B)** A peak with a promoter-like signature (red bar) was located by the transcription start site of *Txnrd1*. The height of the peak was 36 while the input peak was below 16. The peak datasets were displayed with UCSC genome browser [126]. **(C)** The ChIP-seq datasets corresponded to previous ChIP-qPCR results. The graphs represent enrichment levels relative to input from two of the six biological ChIP replicates. qPCR was run with primers of genes *Nqo1* and *Txnrd1*. Both genes were highly enriched in the same NT-NRF2 ChIP samples compared to negative controls such as 66cl4 NRF2 KD samples or normal rabbit IgG-precipitated samples.

3.5 Overlapping findings of ChIP-seq and RNA-seq data

3.5.1 Three known-NRF2-regulated genes identified

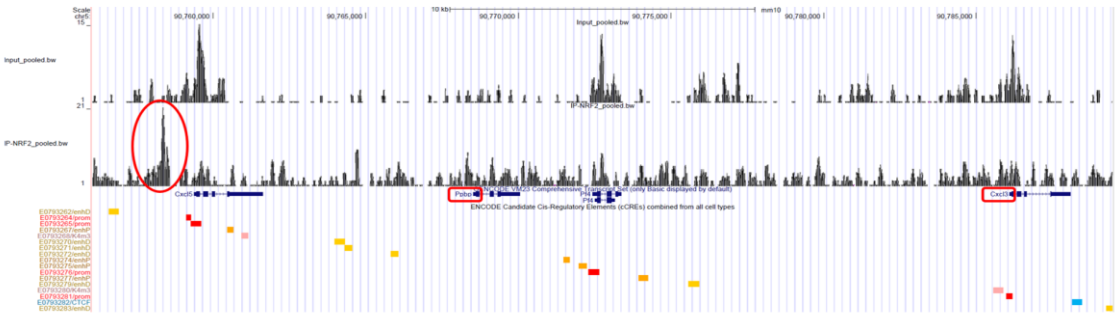
To ensure the identified 35 peaks obtained from ChIP-seq conform with previous findings acquired from the same 66cl4 cells, the list of 40 known-NRF2-modulated genes that were situated in vicinity of the 35 enriched peaks were contrasted against the RNA-seq results. The genes were compared with the list of seven commonly down-regulated genes of 66cl4 NRF2 KD 3B4 and 3B7 cell lines attained from RNA seq (**Supplementary Table 3**). As mentioned earlier, the seven genes were Enrichr-confirmed to be NRF2-regulated [103-105]. The seven genes were thus searched upon the list of 40 genes. Consequently, three genes were spotted to be overlapping (3/7). The genes were *Slc48a1*, *Enah* and *Sim2*. SLC48A1 is a heme transporter that has been reported to be regulated by NRF2 and are recognized to participate in detoxification strategies similarly to other NRF2-induced enzymes [133, 134]. SIM2 is a transcription factor that is known for its vital role in development of the central nervous system while ENAH is an actin-associated protein that supports cell motility and adhesion [135, 136]. However, the two latter proteins have not been extensively investigated on its association to NRF2 in earlier literatures compared to SLC48A1. Taken together, the three NRF2-targeted genes identified in both ChIP-seq and RNA-seq results strengthens the probability of the 35 ChIP-seq peak regions to be NRF2-binding sites in 66cl4 cells.

3.5.2 Two intersecting potential NRF2-regulated genes identified

With the intention to detect potential NRF2-targeted genes overlapping in ChIP-seq and RNA-seq data, the two datasets were again compared. Among the 40 commonly repressed genes from the RNA-seq results, 32 (32/40) have not been identified to be targeted by NRF2 in prior

studies. Thus, the 32 genes were searched upon the NT-NRF2 ChIP dataset. Consequently, gene *Ppbp* was observed approximately 10 kbp distant from a considerable ChIP-seq peak with an evident height of 21 (**Figure 17. A**). This peak was not in the list of 94 significant peaks. However, the low background signals in 66cl4 NT input samples demonstrate the binding of NRF2 to the NT-NRF2 ChIP peak region. Although a 15-heighted peak was found in 66cl4 NT input sample, it was located 1,125 bp downstream from the peak of interest. Interestingly, an additional intersecting gene, *Cxcl3*, was identified roughly 30 kb away from the peak. Unlike *Ppbp*, gene *Cxcl3* was not included within the list of the 40 most transcriptionally-downregulated genes. The gene was even so notably repressed in the 66cl4 NRF2 KD cell lines suggesting the upregulated transcription in 66cl4 NT cells (**Figure 17. B**). *Ppbp* and *Cxcl3* both encode chemokines which act as chemoattractant for neutrophils [127, 137, 138]. Collectively, the two genes observed in RNA-seq and ChIP-seq datasets suggest findings of novel NRF2 target genes in 66cl4 cells that engage in the immune responses.

A



B

	66cl4 NRF2 KD 3B4 / 66cl4 NT		66cl4 NRF2 KD 3B7 / 66cl4 NT		Average	
Gene	Log2(fold-Change)	padj	Log2(fold-Change)	padj	Log2(fold-Change)	padj
Cxcl3	-1.858	0.000532059	-1.944	0.000288662	-1.901	0.0004103605

Figure 17. Two potential NRF2-regulated genes were found in both ChIP-seq and RNA-seq data. **(A)** In the anti-NRF2 antibody EP1808Y-precipitated 66cl4 NT samples (NT-NRF2 ChIP, second lane), genes *Ppbp* and *Cxcl3* were each observed approximately 10 kilobase pairs (kbp) and 30 kbp distant from a considerable ChIP-seq peak with a height of 21. Although a 66cl4 NT input peak with a height of 15 was detected in proximity (first lane), the two peaks were separated with an 1,125 bp gap. This indicates a specific NRF2-binding to the peak region in NT-NRF2 ChIP sample. The peaks are presented with UCSC genome browser [126]. **(B)** Gene *Cxcl3* was not encompassed within the list of the 40 most transcriptionally repressed genes of 66cl4 NRF KD cells in RNA-seq. Nevertheless, it featured statistically significant values suggesting downregulation in NRF2 KD samples. Abbreviation: adjusted P-value (padj).

3.6 Bioinformatic analyses of ChIP-seq peak genes

3.6.1 Both novel and known NRF2-regulated genes are found to participate in a several functions

As a large portion of genes proximate to ChIP-seq peaks are now verified to be regulated by NRF2, the genes were further analyzed to observe additional functions they may be involved in. All 249 genes in vicinity of ChIP peaks were subsequently compared with pathway-gene sets from Enrichr (**Table 1**) [104, 105]. The KEGG 2021 Human and BioPlanet 2019 pathway gene sets confirmed nine genes (*GSTM3*, *GSTM2*, *GSTM1*, *MGST2*, *EPHX1*, *GSTP1*, *HMOX1*, *ABCC2*, *BLVRB*) to participate in xenobiotic metabolism where most of them (*GSTM3*, *GSTM2*,

GSTM1, *MGST2*, *GSTP1*, *GGCT*) are specifically a component of glutathione metabolism pathway [139-142]. Moreover, three genes (*GSTP1*, *ABCC1*, *ABCC2*) were identified to express multi-drug resistance factors, four genes (*HMOX1*, *CSF1*, *IL1R2*, *CD9*) to be involved in IL-6/JAK/STAT3 signaling pathway and seven genes (*PAK6*, *NFAT5*, *CAMK2D*, *VWF*, *MYH1*, *NFKB2*, *ACTG1*) to participate in inflammation mediated by chemokine and cytokine signaling pathway according to pathway gene sets BioCarta 2016, MSigDB Hallmark 2020 and Panther 2016, respectively [143-146]. Among the genes, 13 (12/21) have been identified as NRF2-regulated genes from previous studies, whereas the remaining eight genes (8/21) have not been linked to NRF2 [112-114]. This suggests that the novel NRF2-regulated genes could possibly contribute to major biological functions.

Table 1. Known and potential NRF2-regulated genes are involved in vital biological functions. Some genes in vicinity of the significant ChIP-seq peaks were identified to be components of major pathways according to Enrichr and its constituting genomic studies [104, 105]. The pathways were xenobiotic metabolism, multi-drug resistant factors, IL-6/JAK/STAT3 signaling pathway and inflammation mediated by chemokine and cytokine signaling pathway. Although majority of the genes partaking in the pathways are literature-derived NRF2-regulated genes, eight genes have not been recognized as NRF2-targets to date.

Pathway	Known NRF2-regulated genes	References	Potential NRF2-regulated genes
Xenobiotic metabolism	*GSTM3*, *GSTM1*, *MGST2*, *EPHX1*, *GSTP1*, *HMOX1*, *ABCC2*, *BLVRB*	[139-142, 147]	
Glutathione metabolism	*GSTM3*, *GSTM1*, *MGST2*, *GSTP1*		*GGCT*
Multi-drug resistant factors	*GSTP1*, *ABCC1*, *ABCC2*	[143]	
IL-6/JAK/STAT3 signaling pathway	*HMOX1*, *CD9*	[144]	*CSF1*, *IL1R2*
Inflammation mediated by chemokine and cytokine signaling pathway	*PAK6*, *CAMK2D*	[145, 146]	*NFAT5*, *VWF*, *MYH1*, *NFKB2*, *ACTG1*

3.6.2 Thirty-five genes are highly mutated in clinical metastatic breast cancer samples

To discover the association ChIP-seq peaks may have with metastatic breast cancer, the 249 genes in close distance to ChIP peaks were queried in 16 clinical breast cancer studies under cBioPortal [108, 109]. However, due to the limited number of genes allowed to be analyzed simultaneously on the website, 249 genes were divided into groups and separately examined. Subsequently, 35 genes (35/249) that were found to be altered in more than 3% of patient cohorts were selected for further analyzation. As a result, 36 % (2312) of queried patients were found to have a modification in one or more of the 35 genes (**Supplementary Figure 6**).

Among the 35 genes, 14 genes were each altered in more than 10 % of patient cohorts. Following the analysis, the 14 significantly altered genes were compared with the ENCODE- and ChEA-derived known NRF2-regulated genes to observe an intersection. Only four (4/14; *ENAH*, *EPHX1*, *SRP9*, *TBCE*) were observed to overlap. However, the remaining ten (10/14; *ARID4B*, *B3GLAT2*, *CDC73*, *GLRX2*, *HHIPL2*, *MIA3*, *OTUD6B*, *PIP4P2*, *TAF1A*, *UCHL5*) have not been connected to NRF2 in previous literatures.

In addition, a substantial fraction of patients identified with alterations in one or more of the 35 genes were diagnosed with metastatic breast cancer (**Supplementary Figure 7**). The metastatic breast cancer study from INSERM exhibited genetic alterations in 65.28% (216) of cases [148]. Similar studies from the Metastatic Breast Cancer project data and TCGA each showed alterations in 58.65% (237) and 56.18% (817) cases [108, 109, 149]. Taken together, the statistics suggest a possibility of both known and novel NRF2-regulated genes in 66cl4 NT to assist breast cancer cells in metastatic progression.

3.7 ChIP-qPCR validates the analyzed ChIP-seq datasets

To validate the ChIP-seq datasets, a ChIP-qPCR experiment was essential. For thorough verification, primer sets were generated from ChIP-seq peak regions. From the list of 94 ChIP peaks, peaks located in proximity of known-NRF2-associated genes were chosen to be used for primer production. As these genes are ensured to be NRF2-affiliated by other studies, preparing their primers would enhance the ChIP-qPCR verification method. Among the list of peaks which are presumed to be NRF2-related, the most significant nine peaks were selected based on its $-\log 10\ (Q\ value)$. The nine peaks were proximate to genes *Slc48a1*, *Nbea*, *Krt9*, *Sema3e*, *Hmox1*, *Fth1*, *Slc40a1*, *Abcc1* and *Rpa1*. Thus, primers were generated from DNA sequences of these nine peak regions (**Supplementary Table 2**).

Using the nine prepared primers, ChIP-qPCR was conducted with the remaining ChIP biological replicate samples. The experiment, however, was only performed once due to insufficient volume of residual samples. Identical to the normalization method used in previous qPCR examinations, C_T values of each sample were normalized with adjusted input values to deduce a percent input result [89]. Consequently, 66cl4 NT ChIP samples enriched with anti-NRF2 antibody (EP1808Y) augmented the most considerable amplification levels with an average percent input value of 0.15 (**Figure 18**). However, 66cl4 NT-normal rabbit IgG, 66cl4 NRF2 KD 3B7-EP1808Y and 66cl4 NRF2 KD 3B7-normal rabbit IgG samples yielded an average percent input of 0.0008, 0.0254 and 0.0004, respectively. The clear contrast indicates that EP1808Y-enriched 66cl4 NT samples contained sequences where most of the primers could bind to, in contrary to the other samples. In other words, the results suggests that the ChIP-seq dataset analysis has been effective and hence validates the data.

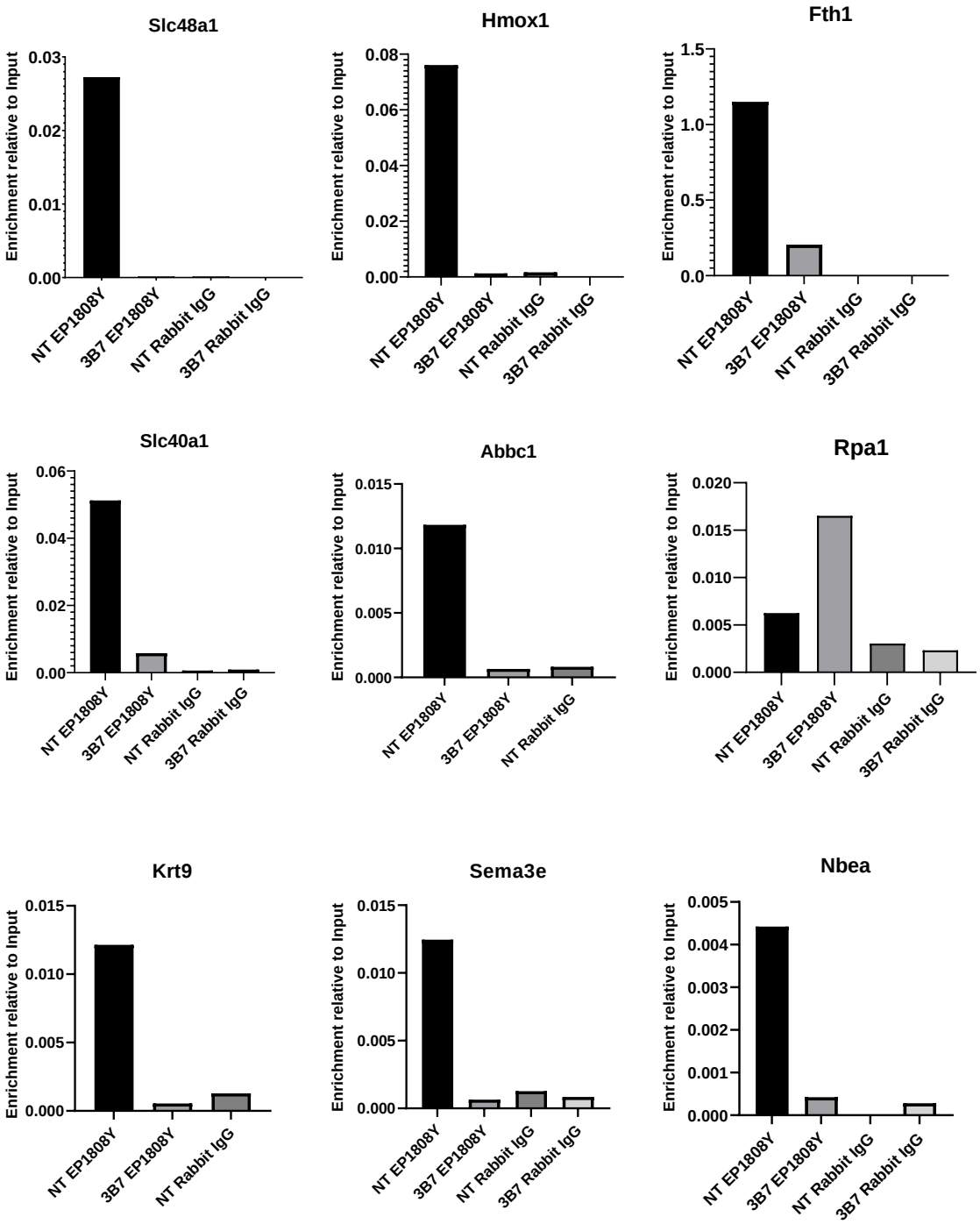


Figure 18. Following ChIP-seq, ChIP-qPCR was conducted which validated the bioinformatically analyzed ChIP-seq dataset. Eight genes (8/9) were highly enriched in NRF2 antibody-enriched 66cl4 NT (NT EP1808Y), while values of 66cl4 NRF2 KD 3B7 (3B7) and normal rabbit IgG (rabbit IgG) ChIP samples were approximately zero. Primer sequences of NRF2-associated genes were selected from ChIP peak regions contiguous to those genes. The DNA sequences of the peak regions were extracted with UCSC genome browser and organized for primer production [126]. The Y-axis indicates percent input enrichment levels normalized by input values [89].

4 Discussion

4.1 Biological findings

NRF2 is a TF that is suggested to be one of the major components of promoting cancer metastasis and has been a focus of research in various types of cancers [14]. However, an understanding of its role in breast cancer is lacking. Thus, this study has performed a global analysis of NRF2 binding sites in 66cl4 murine metastatic breast cancer cells through ChIP-seq. The genome-scale investigation was conducted with 66cl4 cells that constitutively express NRF2 and 66cl4 cells deficient in NRF2. The ChIP method used in this study were optimized for the experimental cell line. Using the tailored method followed by high-throughput sequencing, the results revealed 94 significant ChIP peaks where 59 of them have not been previously identified. Selected peaks were verified using ChIP-qPCR. The results indicate a possible finding of novel NRF2-binding sites in 66cl4 cells.

4.1.1 The 94 significant ChIP-seq peak regions are NRF2-binding sites

The present study identified prominent 94 ChIP-seq peaks in the NRF2 antibody immunoprecipitated-66cl4 NT (NT-NRF2 ChIP) sample (**Supplementary Table 5**) and a solitary peak in the 66cl4 NRF2 KD 3B7 sample (**Supplementary Figure 5**). This is reflective of the ample NRF2 detected in 66cl4 NT cells and the deficient concentration of the TF in 66cl4 NRF2 KD cells according to the Western blotting results (**Figure 4**). Consistently, RNA-seq findings showed downregulation of NRF2-targeted genes in 66cl4 NRF2 KD cell lines when contrasted to the 66cl4 NT cell line (**Supplementary Table 3**). This indicates the mRNA expression of NRF2-modulated genes to be lopsided in 66cl4 NT cells. Furthermore, this is in

agreement with the prior results achieved in our research group where 66cl4 NT cells were found to constitutively express NRF2, while both transcription and translation levels of NRF2 were diminished by approximately 90 % in the 66cl4 NRF2 KD cell lines [61]. Correspondingly, copious quantities of NRF2-targeted DNA fragments would be precipitated in NT-NRF2 IP samples due to the considerable concentration of NRF2 carried in the cell line. Thus, the discrete difference in the number of ChIP-seq peaks between the NT and NRF2 KD cell line suggest the significant peak regions in the NT samples to be NRF2-binding sites.

4.1.2 Genes in vicinity of the significant ChIP-seq peaks are potential NRF2-regulated genes

The NRF2-binding peak regions in NT-NRF2 ChIP samples were located either within or in proximity of 249 genes. A fraction of the genes (40/249) was verified to be targeted by NRF2 in previous literatures (**Figure 14**). The genes were dispersed across 35 peaks which exceeds one-third of all detected peaks. This suggests the remaining genes to be also involved in the NRF2 pathway. The elevated transcriptional expression of known NRF2-modulated genes in NT cells from the RNA-seq results further supports this hypothesis. For instance, putative NRF2-regulated genes *Slc48a1, Enah, Sim2, Alox5ap, Hipk2, Gsta4* and *Tsku* were transcriptionally upregulated in NT cells, where three of the genes (*Slc48a1, Enah, Sim2*) were moreover identified adjacent to significant ChIP-seq peaks. This was validated by the elevated enrichment results of *Slc48a1* in ChIP-qPCR assay of NT-NRF2 ChIP samples (**Figure 18**). This indicates active NRF2/ARE signaling pathway in the NT cell line.

As the genes are regulated by NRF2, the products of the genes were expected to corporate in the detoxification pathway which is a central NRF2-induced mechanism. However, the genes were instead documented to be separately involved in disparate functions (**Table 1**) [133-136,

150]. Intriguingly, these results build on recent evidence that NRF2 is not restricted to a single biological passage but holds a versatile role in diverse regulatory networks of the immune system, inflammation, metabolic reprogramming, iron homeostasis, mitochondrial biogenesis and more [151-153].

However, there are a few limitations in generalizing that all ChIP peak-neighboring genes are targeted by NRF2. The peak genes were recognized upon several factors including whether it was encompassed within a window length of 1 Mbp. Although this was determined on account of the maximal length that distal enhancers can be positioned away from the TSS, this criterion may have inevitably included NRF2-independent genes. Furthermore, in spite of various genes being transcriptionally active in NT cells, it cannot be assumed that all genes are directly regulated by NRF2 as other mediators may additionally be involved in the process. Hence, investigation of NRF2-binding sites in additional metastatic breast cancer cell lines are warranted. Nonetheless, known NRF2-regulated genes implicates the remaining peak genes to be mainly modulated by NRF2. This comprehensive finding reveals potential novel-NRF2-regulated genes in the murine metastatic breast cancer cell line.

4.1.3 Putative and novel NRF2-regulated genes are associated to metastatic breast cancer

The abrogated role of NRF2 in lung cancers or squamous cell carcinomas have been long recognized, but its role in invasive breast cancer has received less attention. Due to limited research on the topic, whether NRF2 promotes or prevents breast cancer malignancy remains to be rather vague. A recent study by Zhou et al. states that hyperactivated NRF2/ARE pathway promotes abnormal ROS regulation which was observed to provoke malignancy in breast carcinoma [154]. An additional study further maintains that NRF2 is a prognostic factor of

breast cancer as NRF2 was found to strongly correlate with invasiveness, recurrence and poor clinical outcomes in breast cancer patients [155]. In accordance with the studies, 35 ChIP-seq peak genes in NT-NRF ChIP samples, comprising both putative and novel NRF2-regulated genes, were found to be considerably modified in breast cancer patients and particularly in metastatic subtypes (**Supplementary Figure 7**). Novel genes such as *TAF1A*, *HHIPL2* and *MIA3* were significantly altered. NRF2-modulated genes identified by RNA-seq exhibited similar results (**Figure 6**). For example, gene *ENAH*, a known NRF2-regulated gene identified in both ChIP- and RNA-seq results, showed substantial genomic alterations in invasive breast cancer [108, 109]. The data contributes to a clearer understanding of specific NRF2-regulated genes that may promote aggressive breast cancer. However, further research is needed to establish the network of how the NRF2-regulated genes advances the metastatic features of breast cancer.

On the other hand, Giudice et al. asserts that NRF2 activation prevents malignant development of estrogen receptor-mediated breast tumor by redox cycling of reactive estrogen quinones and ROS [156]. This suggests the contradicting role of NRF2 in inhibiting breast oncogenesis in normal cells but promoting metastasis in breast cancer cells. This has been reported likewise in other cancers [157]. However, what biological mechanism shifts the opposing status remains to be clarified [120]. Hence, further studies should conduct extensive identification of genomic NRF2-regulations to understand both its role of repressing and promoting breast cancer.

4.1.4 NRF2-regulated genes of 66cl4 cells express drug metabolizing enzymes

The present study has suggested that genes in vicinity to the 94 significant ChIP peak regions are known and potential NRF2-regulated genes that are involved in major biological functions (**Table 1**) [104, 105]. The most notable was the xenobiotic metabolism pathway, also known

as drug metabolism which is a major NRF2-induced function. Drug metabolism refers to enzymatic modification of foreign compounds which shows resistance against oxidants or electrophiles [158]. The pathway is divided into three phases (phase I; modification, phase II; conjugation and phase III; excretion) [159, 160].

In this study, the network of genes encoding components of the drug metabolism pathway were *Ephx1*, *Hmox1*, *Gstm1*, *Gstm2*, *Gstm3*, *Gstp1*, *Gstp2*, *Mgst2*, *Abcc1*, *Abcc2*, and *Ggct*. EPHX1 and HMOX1 are enzymes that catalyze phase I reactions, while GSTMs, GSTPs and MGST2 are glutathione S-transferases (GSTs) which are essential enzymes of phase II conjugation reaction [11]. The pathway can be initiated by oxidative stress. GSTs can then conjugate glutathione to ROS and electrophiles for detoxification. The conjugated metabolites are equipped for transportation through efflux transporters such as ABCC1 and ABCC2 which are multidrug resistance proteins [159, 161].

The known NRF2-modulated drug metabolizing enzymes augment protective effects in healthy cells. However, they have also been reported to support invasiveness of cancer cells [44, 45]. NRF2-addicted cancer cells upregulate expression of the enzymes which enables cells to evade chemotherapeutic drugs by glutathione conjugation and subsequent excretion [155, 162]. NRF2-mediated resistance to chemotherapeutic drugs have been observed in different cancers. Among them are aggressive breast cancer types such as triple negative breast cancer (TNBC). In TNBC, NRF2 was found to confer resistance to cisplatin, doxorubicin and mitoxantrone which are all drugs commonly used for metastatic breast cancer. [155, 163, 164].

In addition, a potential NRF2-targeted gene involved in drug metabolism, *Ggct*, was identified in our analyses. GGCT is a crucial enzyme participating in glutathione metabolism [165]. It

has besides been linked to poor clinical outcome in multiple carcinomas including breast cancer [166-169].

The identification of putative and potential NRF2-regulated genes participating in drug metabolism reinforce the connection between NRF2 and detoxification mechanisms. This is consistent with previous findings and hence strengthens the significance of the detected ChIP-seq peaks. The ChIP-seq analyses further suggested a possibility of the pathway enzymes to support tumorigenesis through aberrated NRF2 activity. It cannot be denied, however, that a detailed description is needed on how upregulated NRF2-drug metabolism provokes metastatic breast cancer. Nonetheless, the NRF2-enriched AREs from 66cl4 NT cells imply a close correlation between NRF2-regulated antioxidant proteins and metastatic breast cancer.

4.1.5 Role of NRF2 in the tumour microenvironment of breast cancer

A tumor microenvironment (TME) refers to the tumor site combined with a variety of factors such as the tumor cells itself, infiltrated immune cells, stromal cells, blood vessels and more. A common role of these components is stimulating tumor progression while suppressing anti-tumor functions [170]. Cells and proteins of the immune system have been a specific focus in the field as dysregulated inflammatory mechanisms are known to expose the inflamed site with the risk of neoplastic transformation through excessive production of ROS which can damage DNA [171]. Despite active research on TME, its underlying mechanisms remain to be further investigated in various types of tumors including breast cancer. Thus, it is of interest to identify how NRF2 may contribute to the TME in aggressive breast cancer.

From this study, the ChIP-seq results revealed peak enrichments in proximity to notable TME-promoting components which were colony stimulating factor (*CSF1*), C-X-C motif chemokine

ligand 3 (*CXCL3*), pro-platelet basic protein (*PPBP*), nuclear factor of activated T cells 5 (*NFAT5*), nuclear factor kappa B subunit 2 (*NFKB2*) and Interleukin 1 Receptor Type 2 (*IL1R2*). These genes are potential NRF2-regulated genes.

CSF1 are cytokines required for differentiation of M2 macrophages which are also known as tumor-associated macrophages (TAMs) [172]. TAMs are major inflammatory players of TME promoting angiogenesis and proliferation [171]. CSF1 also signals TAM to express cytokines and proteases such as interleukin-10 (IL-10) and matrix metalloproteinases to stimulate growth and eventually metastasis [173]. Thus, CSF1 has been considered to correlate with poor prognosis which is also the case in breast cancer. CSF1 has been heavily found in metastatic breast tumors along with infiltrated TAMs and IL-10 [173, 174].

Genes *CXCL3* and *PPBP* are also associated to the TME in metastatic breast cancer. CXCL3 and PPBP are both chemokines that potently attract and activate neutrophils by binding to CXC receptor 2 [127, 138, 175]. The induced neutrophils are the first responders of eliminating microbial pathogens and repairing damaged tissue. However, dysregulated chemokine-neutrophil pathway can inversely promote TME progression. Cancer cells diligently recruit neutrophils by secreting its cognate chemokines. Neutrophil infiltration in tumor subsequently promote metastasis by producing ROS, angiogenic factors and inhibiting anti-tumor immune responses [176]. Overexpression of the two chemokines have been observed in multiple cancers. Elevated levels of PPBP were found to heighten the MAPK and PI3K/AKT/mTOR pathway and have thus been pointed as a potential biomarker [177, 178]. CXCL3 is also highly expressed in cancers including metastatic breast cancer [179]. This is consistent with previous findings from our group which identified increased transcript and protein levels of CXCL3 and large population of neutrophils in 66cl4 cells [180].

In addition, *NFAT5* and *NFKB2* are potential NRF2-targeted genes encoding TFs that take vital

roles in immune response [145, 146, 181]. NFAT5 induces transcription of proinflammatory cytokines in T cells, such as TNF-α [182]. TNF-α is a constituent of TME with a pro-tumorigenic role [183]. Thereby, increased expression of NFAT5 is predictably shown to advance metastases and was further suggested as a biomarker for inflammatory breast cancer [184, 185]. The second prospective NRF2-modulated TF, NFKB2, is a subunit of NF-κB and is a main component of the NF-κB noncanonical pathway. This pathway promotes cell proliferation by enhancing and inhibiting expressions of anti-apoptotic and apoptotic genes, respectively. Overactivation of this pathway have accordingly been discovered to support ovarian cancer, glioma, breast cancers and more [186].

Lastly, IL-1R2 is another protein possibly induced by NRF2. IL-1R2 is a negative receptor of a proinflammatory cytokine interleukin-1 (IL-1) [187, 188]. Despite its IL-1-neutralizing role, IL-1R2 is found to be highly expressed on regulatory T cells (Treg) in several types of cancers including breast tumors [189, 190]. This is supported by findings of Tregs stimulating aggressive malignancies in breast cancer [191]. Taken together, this may indicate that NRF2 contributes to development of aggressive breast cancer through regulating cytokine signaling pathways. However, this cannot be quickly corroborated as more research is required of the mentioned TME-associated components. Nevertheless, similar findings of the six proteins in previous breast cancer research and in the current 66cl4 cell line study suggest the involvement of NRF2 in TME of breast cancers.

4.1.6 Function of NRF2 in heme homeostasis

Senescent red blood cells are degraded in the phagolysosome of macrophages. Heme is then transported into the cytosol via a heme transporter called HRG1 which is thereafter detoxified by HMOX1 [192]. This pathway is essential in sustaining heme levels. Thus, HMOX1 and

HRG1 are required for maintaining heme homeostasis [193, 194]. In this study, genes encoding HRG1 (*Slc48a1*) and HMOX1 (*Hmox1*) were identified. *Slc48a1* was found in both RNA-seq and ChIP-seq results. This implies a possible role NRF2 may take in heme homeostasis.

Slc48a1 and *Hmox1* are known NRF2-targeted genes and have been addressed to have a strong expressional correlation with each other which is in line with our study [195]. This was validated in the ChIP-qPCR experiment where both genes were highly enriched in NT-NRF2 ChIP samples (**Figure 18**). *Slc48a1* has also been linked to metastatic property of breast cancer. HRG-1 was found to facilitate growth and invasion of the carcinoma [196]. Although it cannot be generalized that NRF2-induced *Slc48a1*-upregulation leads to metastasis, findings of the genes in the present study suggest a possible link between NRF2 and heme homeostasis in aggressive breast cancer.

4.1.7 Majority of NRF2 binding sites are located in intronic and intergenic regions

It is common knowledge that TFs bind to promoters adjacent to TSSs or regulatory sequences such as enhancers which can be located nearby or remotely from downstream target genes [197]. The present study is in line with the expectations where more than a half of the significant ChIP peaks were in sequences with enhancer-like signatures. However, there were 44 novel NRF2 peaks that were not characterized with any enhancer or promoter signatures but were located in intronic and intergenic regions. Although NRF2 binding of intronic or intergenic regions were unforeseen, interaction with the sites were considered universal for numerous TFs. In a recent ENCODE project, DNase I hypersensitive sites (DHS) analysis was conducted with 125 human cell types which revealed that 95 % of DHSs were found in intronic and intergenic regions whereas the remaining 5 % represented TSSs [198]. Another study stated introns to be major regulatory sites for gene expression rather than promoters and that

expression of intron-regulated genes are highly expressed in most tissues [199]. Specific findings of NRF2 binding locations were also established. In *ABCC3* gene, NRF2 was found to interact with two AREs, both located in an intron, in order to induce expression of the gene [200]. This supports the detection of novel NRF2 peaks positioned within those regions in this study. Furthermore, it could be inferred that majority of NRF2 binding-AREs that are yet to be discovered may also be located in intronic or intergenic areas, based on the findings.

4.2 Methodological considerations

4.2.1 Optimal DNA fragment size suiting both ChIP-qPCR and high-throughput sequencing

qPCR and sequencing technologies each require varying sizes of DNA fragments. For instance, it is optimal for template sizes to be less than 300 bp for SYBR Green assays [201], whereas Illumina sequencing recommend lengths ranging from 350 to 550 bp depending on type of kit and instrument used [202]. Fragments that are too short can induce poor sequencing quality as the instrument will continue to read the adapter sequences if the insert size is shorter than the sequencing read length [203]. Inversely, there has been reports that fragments longer than 500 bp can induce poor base qualities and increased errors [204]. Finding a settled size can be thus challenging. Despite different ChIP-seq studies recommending varying ChIP DNA lengths, most fall within a range of 100 – 600 bp [66, 69, 205]. As ChIP-seq studies are optimized for both sequencing and PCR methods, we adapted our aimed fragment length following their recommendations but further adjusting based on SYBR Green assay and Illumina guidelines which led us to a length of 200 – 500 bp.

4.2.2 Adequate cell density required for sonication

Sonication is a widely used technique to shear DNA to optimal lengths for ChIP-seq [69]. The method was also used for this study which successfully generated DNA lengths in targeted range of 200 – 500 bp for each biological replicate. However, yielding an optimal size were in times challenging. Following electrophoresis, bands from input samples were on a few occasions located above 500 bp ladder band. It was important to unravel the cause as both under- and over-sheared DNA fragments will directly affect the quality of ChIP-seq NGS data [91, 205]. After several experiments, it was inferred that it may be caused by excessive cell concentration. Each cell line was plated in four dishes and harvested together. Following sonication, a delayed electrophoretic migration was observed for cells that were incubated for 60 hours before harvest. Contrarily, no band retardation was noticed in samples that were incubated for less than 48 hours. The Bioruptor Pico guidelines from Diagenode stated that high sample viscosity leads to inefficient DNA shearing and thus advised a cell density of 1,000,000 – 3,000,000 cells per 100 µl for good fragmentation [206]. Despite their findings being in line with our conjecture, cell numbers within the range till showed dense concentrations leading to unsatisfying results. Input samples of cell line 66cl4 NT, 66cl4 NRF2 KD 3B7, 66cl4 NRF2 KD 3B4 respectively had a cell number of 1,180,000, 1,980,000 and 1,000,000 per 100 µl. However, when cell densities were decreased to 204,000, 242,000 and 140,000 cells, sonication efficiency was significantly improved. Although preparing samples with low cell density is crucial for optimal DNA fragmentation, the finding indicates that the required number of cells may vary considerably depending on the type of cell line.

4.2.3 An alternative DNA fragmentation method

Although the ChIP-seq method has continued to advance ever since its first development in 2007, most of the procedure has maintained the conventional workflow [73]. However, some steps differ depending on the study. One of those steps is the chromatin fragmentation method which have two available options: sonication and/or enzymatic digestion. Each are used mainly for cross-linked ChIP (X-ChIP) and native ChIP (N-ChIP), respectively. N-ChIP refers to ChIP without the crosslinking step and are generally used for studying histone or nucleosome positions [69]. For N-ChIP, cells are treated with MNase as the enzymes produce high-resolution results while the sonication method can disrupt the histone association to DNA [205, 207]. On the other hand, X-ChIP, ChIP integrated with crosslinking step, is used for sequence-specific binding proteins such as TFs. Contrarily to N-ChIP, opinions differ on what type of shearing method is better for X-ChIP. Some state mechanical shearing techniques are more suitable since MNase may digest the TF-binding site [205]. Others refute that MNase inversely manages to provide precise TF-occupied regions with higher resolution by shearing the DNA fragment into smaller lengths which shows more accurate genomic presentation of where the TF binds to [208]. The present study has, however, used the sonication method as many NRF2 ChIP-seq studies have employed the same technique which generated meaningful results [22, 87, 209]. Despite the controversy of which fragmentation method ameliorates the identification of TF-binding sites, it cannot be denied that the conventional method needs to be further optimized for a higher resolution as TFs only binds to an approximate DNA length of 5 – 15 bp [210].

4.2.4 Conservative ChIP-seq peak filtering may have excluded well-known NRF2-regulated genes from list of significant peaks

Nqo1 and *Txnrd1* are considered representative NRF2-regulated genes as their expression have been reported to be upregulated by NRF2 in numerous studies [87, 211, 212]. Despite the genes being detected in the ChIP-seq dataset, the genes were not detected in proximity to any of the 94 significant ChIP peaks. This may be due to conservative criteria established by BioCore for filtering significant peaks from the rest of the peaks. For example, the distinct peak found by the TSS of *Nqo1* is speculated to have been considered algorithmically insignificant due to a 13-sized peak in the NT-input sample which was located 250 bp away from the *Nqo1*-peak region (**Figure 16. A**). This raises a possibility of novel NRF2-regulated genes being disregarded from analyses. However, lowering the bar may start to encompass insignificant peaks that were possibly caused by noise and other variables and hinder identification of NRF2-binding sites. Thus, overlooking a few peak regions may be inevitable to achieve results with maximum signal-to-noise ratio.

4.2.5 Specificity of NRF2 antibodies

The quality of a ChIP-seq data is heavily dependent on the sensitivity and specificity of the antibody that is used to pull down the protein-DNA complex of interest [70]. Antibodies with high sensitivity and specificity will provide clear results of distinctly enriched peaks and low background levels [69]. Accordingly, it is important to predict possible outcomes an ineffectual antibody can bring in order to select one with high quality. Defective antibodies can potentially cross react to other chromosomal proteins or have poor reactivity against its intended target. These factors can respectively bring substantial background noises and negligible enrichments in the region of interest. Fortunately, this can be avoided by carefully selecting the adequate

antibody using PCR assays following ChIP. In ChIP-PCR, several specific loci with control regions can be tested using cell lines including knockout or knockdown models. Antibodies producing elevated signals in positive controls of wild type cells, but ground-level signals in knockdown cells and negative control regions addresses the high sensitivity and minimal cross-reactivity [70, 205].

This study selected recombinant anti-NRF2 antibody (EP1808Y) (Abcam) on account of its high specificity towards NRF2 in 66cl4 murine metastatic breast cancer cells compared to NRF2 (D1Z9C) rabbit mAb (Cell Signaling Technology). EP180Y induced elevated signals in positive control regions (*Nqo1*, *Txnrd1*) of 66cl4 NT cells and the opposite in 66cl4 NRF2 KD 3B7 cells. However, some studies have reversely shown high specificity of D1Z9C to both human and mouse NRF2 [110, 213, 214]. Kemmerer et al. compares four types of NRF2-targeting antibodies including EP1808Y and D1Z9C and states the latter antibody to have the highest sensitivity [110]. Huppke et al. additionally demonstrates high sensitivity and specificity of D1Z9C against NRF2 when cells were treated with NRF2 inducers such as sulforaphane [214]. Taken together, although selecting an antibody of high quality is important, specificity of primary antibodies against NRF2 seem to vary based on the types of cells, usage of an inducer and the overall method utilized in the study. Therefore, it is crucial to take all experimental components into consideration when choosing the adequate NRF2 antibody.

4.2.6 Usage of control samples

Although ChIP-seq is known to read TF-binding sites with the highest resolution among other techniques, it can also be interrupted by background noises due to several artifacts [69, 205]. Firstly, open chromatin regions tend to be sheared more easily than heterochromatin regions. This produces more protein-DNA complexes from the open structures which leads to

unbalanced number of clustered reads. This is referred as a sonication bias [215]. Additionally, cross reactivity of antibodies can occur and affect the results. Some genomic regions have more regulatory proteins and hence ample protein-DNA complexes. Despite the primary antibodies being designed to bind to the TF of interest, it may also bind to other proteins with similar epitopes [205, 215]. As a result, genomic regions with concentrated regulatory proteins generate higher enrichment than sites with less proteins [215]. To avoid these technical and biological biases, it is essential to include controls in the ChIP-seq study.

There are mainly three types of controls that can be used which are the input sample, nonspecific IP sample and mock IP sample [69]. Nonspecific IP samples and mock IP samples are both negative controls. Nonspecific IP samples have chromatins precipitated with a nonspecific antibody, such as normal rabbit IgG, whereas mock IP samples are prepared by conducting an antibody-free IP [69]. However, nonspecific IP controls are more commonly used as mock IP samples precipitate only a minimal amount of DNA which thereby produces inconsistent results [69, 215]. Mock IP sample was not incorporated to the present study due to this reason. Input control, on the other hand, are a portion of the samples that have been separated following DNA sonication and thus have not been immunoprecipitated. These controls are crucial in identifying a high-confidence set of TF-binding sites [87].

Input controls can correct the sonication bias. Since input DNA have also been sheared under the same condition, unevenly enriched genomic regions will be shown in their datasets as well. Thus, overrepresentation of these sites can be removed by comparing the targeted TF-IP sample to this control. This typically occurs in proximal enhancer sites or promoters where the chromatin is open and susceptible for fragmentation [87]. For instance, the input control of 66cl4 NT cells from this study displayed higher ChIP-seq peaks at the TSSs of genes *Hmox1* and *Gstm1* in contrast to the rest of the background (**Figure 15**. **A, B**). This facilitates

calculating the overrepresentation of ChIP peaks identified in the regions of 66cl4 NT-NRF2 ChIP samples.

Furthermore, nonspecific IP controls can amend ChIP-seq data affected by antibody cross-reactivity [70, 215]. Nonspecific antibodies, namely normal rabbit IgG from Cell Signaling Technology, cannot specifically interact with any antigens within the sample but may nonspecifically bind to irrelevant proteins much like the NRF2-targeting antibodies can [70, 83]. Consequently, spurious genomic sites caused by abundant regulatory proteins in 66cl4 NT-NRF2 ChIP-seq results can be removed by contrasting the peaks with nonspecific IP data. Using both input and nonspecific IP controls in this study has enabled the production of meaningful results in determining significantly enriched NRF2-binding sites.

4.2.7 Varying molecular weight of NRF2

Although the manufacturer of the NRF2 (D1Z9C) rabbit mAb (Cell Signaling Technology) states the NRF2 molecular weight to range between 97-100 kDa [216], the detected 75 kDa band in this study was considered as NRF2 since the size seems to vary depending on the types of cell lines and SDS gels used (**Supplementary Figure 1**) [110, 217]. While some studies detect molecular weight of NRF2 to be ~95-110 kDa [22, 63, 217], many other groups have found the size to be 68-80 kDa which conforms with our result [110, 218, 219]. A study from Kemmerer et al. further presented an NRF2 band of ~80 kDa when electrophoresed in the 4-12% gradient Bis-Tris SDS PAGE gel but a size above 100 kDa for NRF2 that has migrated in 10% Tris-glycine SDS gels [110]. This strengthens the finding of 75 kDa-sized NRF2 which was electrophoresed in 4-12% Bis-Tris gels.

5 Conclusion and Future prospects

This study aimed to perform global profiling to investigate NRF2-binding sites in 66cl4 metastatic breast cancer cell line. Based on the combined analyses using an optimized ChIP method, massive parallel sequencing and bioinformatic interpretation, it can be concluded that 94 prominent NRF2-binding regions were identified. The data presented 94 significant ChIP-seq clusters vicinal to known NRF2-regulated genes in 66cl4 NT cells enriched by NRF2 antibodies. ChIP-qPCR validated the quality of the results. On the contrary, the 66cl4 NRF2 KD cells exhibited a single significant ChIP-seq peak indicating the specific binding of NRF2 in 66cl4 NT cells. This was supported by corresponding RNA-seq data. Enrichr- and cBioPortal-mediated research further suggested NRF2 to contribute to extensive biological pathways in metastatic breast cancer.

This study expands current knowledge of NRF2 in metastatic breast cancer. Potentially novel NRF2-regulated genes were highly mutated in aggressive breast cancer biopsies suggesting a contribution of the genes to metastatic abilities. However, this requires further verification. Repetition of the used method in additional metastatic breast cancer cell lines will build generalizability of the results.

To the best of our knowledge, the present study is the first to conduct ChIP-seq on a metastatic breast cancer cell line. ChIP-seq has ameliorated the identification of novel binding sites. However, to better understand the precise NRF2 binding segments of these potential NRF2-regulated genes in metastatic breast cancer cells, future studies could practice reporter gene assays. Partial regions of the NRF2-binding sites can be cloned into a vector contiguously to a reporter gene. Once the vector is transfected back into the cell, the NRF2-induced expressional activity of the reporter proteins will indicate whether the transfected partial promoter or enhancer is a core region for interacting with NRF2 and subsequently initiating transcription.

Furthermore, this research has briefly introduced major biological networks that the potential NRF2-targeted genes partake in, such as the tumor microenvironment. This shows that NRF2 may have a greater role in regulating critical physiological conditions. Thus, the presented genes can be utilized to enrich the comprehension on the underlying NRF2-mechanisms in metastatic breast cancer. A full discovery of the NRF2-regulatory system can be uncovered by comparative analyses employing ChIP-seq. Contrasting NRF2-binding sites of healthy breast cells to its malignant type may help deciphering the process of how the genes promote tumorigenesis and metastases. Differentiating tissue-specific NRF2-binding sites using ChIP-seq data collected from diverse cancer studies can also assist in detailing functional NRF2-regulated genes in breast cancer. An exhaustive understanding of the NRF2-directed genomic mechanisms will be a crucial step in developing translational research of metastatic breast cancer.

6 References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* 2021, 71(3):209-249.
2. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN *et al*: The repertoire of mutational signatures in human cancer. *Nature* 2020, 578(7793):94-101.
3. Roychowdhury S, Chinnaiyan AM: Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin* 2016, 66(1):75-88.
4. Turajlic S, Sottoriva A, Graham T, Swanton C: Resolving genetic heterogeneity in cancer. *Nat Rev Genet* 2019, 20(7):404-416.
5. Finishing the euchromatic sequence of the human genome. *Nature* 2004, 431(7011):931-945.
6. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, Perry MD, Nahal-Bose HK, Ouellette BFF, Li CH *et al*: Pan-cancer analysis of whole genomes. *Nature* 2020, 578(7793):82-93.
7. Nogrady B: How cancer genomics is transforming diagnosis and treatment. *Nature Outlook: Cancer diagnosis* 2020, 579:S10-S11.
8. Bushweller JH: Targeting transcription factors in cancer — from undruggable to reality. *Nature Reviews Cancer* 2019, 19(11):611-624.
9. Lambert M, Jambon S, Depauw S, David-Cordonnier MH: Targeting Transcription Factors for Cancer Treatment. *Molecules* 2018, 23(6).
10. Moi P, Chan K, Asunis I, Cao A, Kan YW: Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region. *Proc Natl Acad Sci U S A* 1994, 91(21):9926-9930.
11. Zhang Y: Phase II Enzymes. In: *Encyclopedia of Cancer*. Edited by Schwab M. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011: 2853-2855.
12. Jancova P, Anzenbacher P, Anzenbacherova E: Phase II drug metabolizing enzymes. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2010, 154(2):103-116.
13. Taguchi K, Yamamoto M: The KEAP1–NRF2 System in Cancer. *Frontiers in Oncology* 2017, 7(85).
14. Namani A, Liu K, Wang S, Zhou X, Liao Y, Wang H, Wang XJ, Tang X: Genome-wide global identification of NRF2 binding sites in A549 non-small cell lung cancer cells by ChIP-Seq reveals NRF2 regulation of genes involved in focal adhesion pathways. *Aging (Albany NY)* 2019, 11(24):12600-12623.
15. Hirotsu Y, Katsuoka F, Funayama R, Nagashima T, Nishida Y, Nakayama K, Douglas Engel J, Yamamoto M: Nrf2–MafG heterodimers contribute globally to antioxidant and metabolic networks. *Nucleic Acids Research* 2012, 40(20):10228-10239.
16. Nguyen T, Nioi P, Pickett CB: The Nrf2-antioxidant response element signaling

pathway and its activation by oxidative stress. *J Biol Chem* 2009, 284(20):13291-13295.
17. Suzuki T, Yamamoto M: Molecular basis of the Keap1-Nrf2 system. *Free Radic Biol Med* 2015, 88(Pt B):93-100.
18. DeNicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, Mangal D, Yu KH, Yeo CJ, Calhoun ES *et al*: Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* 2011, 475(7354):106-109.
19. Ito A, Shimazu T, Maeda S, Shah AA, Tsunoda T, Iemura S, Natsume T, Suzuki T, Motohashi H, Yamamoto M *et al*: The subcellular localization and activity of cortactin is regulated by acetylation and interaction with Keap1. *Sci Signal* 2015, 8(404):ra120.
20. Wu WL, Papagiannakopoulos T: The Pleiotropic Role of the KEAP1/NRF2 Pathway in Cancer. *Annual Review of Cancer Biology* 2020, 4(1):413-435.
21. Tong KI, Kobayashi A, Katsuoka F, Yamamoto M: Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol Chem* 2006, 387(10-11):1311-1320.
22. Kobayashi EH, Suzuki T, Funayama R, Nagashima T, Hayashi M, Sekine H, Tanaka N, Moriguchi T, Motohashi H, Nakayama K *et al*: Nrf2 suppresses macrophage inflammatory response by blocking proinflammatory cytokine transcription. *Nat Commun* 2016, 7:11624.
23. Tonelli C, Chio IIC, Tuveson DA: Transcriptional Regulation by Nrf2. *Antioxid Redox Signal* 2018, 29(17):1727-1745.
24. Taguchi K, Yamamoto M: The KEAP1-NRF2 System as a Molecular Target of Cancer Treatment. *Cancers (Basel)* 2020, 13(1).
25. Taguchi K, Motohashi H, Yamamoto M: Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. *Genes Cells* 2011, 16(2):123-140.
26. Nam LB, Keum YS: Binding partners of NRF2: Functions and regulatory mechanisms. *Arch Biochem Biophys* 2019, 678:108184.
27. Rachakonda G, Xiong Y, Sekhar KR, Stamer SL, Liebler DC, Freeman ML: Covalent modification at Cys151 dissociates the electrophile sensor Keap1 from the ubiquitin ligase CUL3. *Chem Res Toxicol* 2008, 21(3):705-710.
28. Zhou XL, Zhu CY, Wu ZG, Guo X, Zou W: The oncoprotein HBXIP competitively binds KEAP1 to activate NRF2 and enhance breast cancer cell growth and metastasis. *Oncogene* 2019, 38(21):4028-4046.
29. Sun Z, Chin YE, Zhang DD: Acetylation of Nrf2 by p300/CBP augments promoter-specific DNA binding of Nrf2 during the antioxidant response. *Mol Cell Biol* 2009, 29(10):2658-2672.
30. Sekine H, Okazaki K, Ota N, Shima H, Katoh Y, Suzuki N, Igarashi K, Ito M, Motohashi H, Yamamoto M: The Mediator Subunit MED16 Transduces NRF2-Activating Signals into Antioxidant Gene Expression. *Mol Cell Biol* 2016, 36(3):407-420.
31. Sun Z, Wu T, Zhao F, Lau A, Birch CM, Zhang DD: KPNA6 (Importin {alpha}7)-mediated nuclear import of Keap1 represses the Nrf2-dependent antioxidant response. *Mol Cell Biol* 2011, 31(9):1800-1811.

32. Kobayashi A, Kang MI, Okawa H, Ohtsuji M, Zenke Y, Chiba T, Igarashi K, Yamamoto M: Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. *Mol Cell Biol* 2004, 24(16):7130-7139.
33. Valko M, Leibfritz D, Moncol J, Cronin MT, Mazur M, Telser J: Free radicals and antioxidants in normal physiological functions and human disease. *Int J Biochem Cell Biol* 2007, 39(1):44-84.
34. Johnson NM, Egner PA, Baxter VK, Sporn MB, Wible RS, Sutter TR, Groopman JD, Kensler TW, Roebuck BD: Complete protection against aflatoxin B(1)-induced liver cancer with a triterpenoid: DNA adduct dosimetry, molecular signature, and genotoxicity threshold. *Cancer Prev Res (Phila)* 2014, 7(7):658-665.
35. Iida K, Itoh K, Maher JM, Kumagai Y, Oyasu R, Mori Y, Shimazui T, Akaza H, Yamamoto M: Nrf2 and p53 cooperatively protect against BBN-induced urinary bladder carcinogenesis. *Carcinogenesis* 2007, 28(11):2398-2403.
36. Long M, Tao S, Rojo de la Vega M, Jiang T, Wen Q, Park SL, Zhang DD, Wondrak GT: Nrf2-dependent suppression of azoxymethane/dextran sulfate sodium-induced colon carcinogenesis by the cinnamon-derived dietary factor cinnamaldehyde. *Cancer Prev Res (Phila)* 2015, 8(5):444-454.
37. Kim EH, Deng C, Sporn MB, Royce DB, Risingsong R, Williams CR, Liby KT: CDDO-methyl ester delays breast cancer development in BRCA1-mutated mice. *Cancer Prev Res (Phila)* 2012, 5(1):89-97.
38. Yoh K, Itoh K, Enomoto A, Hirayama A, Yamaguchi N, Kobayashi M, Morito N, Koyama A, Yamamoto M, Takahashi S: Nrf2-deficient female mice develop lupus-like autoimmune nephritis. *Kidney Int* 2001, 60(4):1343-1353.
39. Ma Q, Battelli L, Hubbs AF: Multiorgan autoimmune inflammation, enhanced lymphoproliferation, and impaired homeostasis of reactive oxygen species in mice lacking the antioxidant-activated transcription factor Nrf2. *Am J Pathol* 2006, 168(6):1960-1974.
40. Frohlich DA, McCabe MT, Arnold RS, Day ML: The role of Nrf2 in increased reactive oxygen species and DNA damage in prostate tumorigenesis. *Oncogene* 2008, 27(31):4353-4362.
41. Kitamura Y, Umemura T, Kanki K, Kodama Y, Kitamoto S, Saito K, Itoh K, Yamamoto M, Masegi T, Nishikawa A *et al*: Increased susceptibility to hepatocarcinogenicity of Nrf2-deficient mice exposed to 2-amino-3-methylimidazo[4,5-f]quinoline. *Cancer Sci* 2007, 98(1):19-24.
42. Ramos-Gomez M, Dolan PM, Itoh K, Yamamoto M, Kensler TW: Interactive effects of nrf2 genotype and oltipraz on benzo[a]pyrene-DNA adducts and tumor yield in mice. *Carcinogenesis* 2003, 24(3):461-467.
43. Enomoto A, Itoh K, Nagayoshi E, Haruta J, Kimura T, O'Connor T, Harada T, Yamamoto M: High sensitivity of Nrf2 knockout mice to acetaminophen hepatotoxicity associated with decreased expression of ARE-regulated drug metabolizing enzymes and antioxidant genes. *Toxicol Sci* 2001, 59(1):169-177.

44. Padmanabhan B, Tong KI, Ohta T, Nakamura Y, Scharlock M, Ohtsuji M, Kang MI, Kobayashi A, Yokoyama S, Yamamoto M: Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell* 2006, 21(5):689-700.
45. Leinonen HM, Kansanen E, Pölönen P, Heinäniemi M, Levonen AL: Dysregulation of the Keap1-Nrf2 pathway in cancer. *Biochem Soc Trans* 2015, 43(4):645-649.
46. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012, 489(7417):519-525.
47. Kim YR, Oh JE, Kim MS, Kang MR, Park SW, Han JY, Eom HS, Yoo NJ, Lee SH: Oncogenic NRF2 mutations in squamous cell carcinomas of oesophagus and skin. *J Pathol* 2010, 220(4):446-451.
48. Goldstein LD, Lee J, Gnad F, Klijn C, Schaub A, Reeder J, Daemen A, Bakalarski CE, Holcomb T, Shames DS *et al*: Recurrent Loss of NFE2L2 Exon 2 Is a Mechanism for Nrf2 Pathway Activation in Human Cancers. *Cell Rep* 2016, 16(10):2605-2617.
49. Wang R, An J, Ji F, Jiao H, Sun H, Zhou D: Hypermethylation of the Keap1 gene in human lung cancer cell lines and lung cancer tissues. *Biochem Biophys Res Commun* 2008, 373(1):151-154.
50. Yoo NJ, Kim HR, Kim YR, An CH, Lee SH: Somatic mutations of the KEAP1 gene in common solid cancers. *Histopathology* 2012, 60(6):943-952.
51. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M *et al*: Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genetics* 2015, 47(2):106-114.
52. Yamamoto M, Kensler TW, Motohashi H: The KEAP1-NRF2 System: a Thiol-Based Sensor-Effector Apparatus for Maintaining Redox Homeostasis. *Physiol Rev* 2018, 98(3):1169-1203.
53. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S *et al*: Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* 2018, 173(2):321-337.e310.
54. Sánchez-Martín P, Saito T, Komatsu M: p62/SQSTM1: 'Jack of all trades' in health and cancer. *Febs j* 2019, 286(1):8-23.
55. Komatsu M, Kurokawa H, Waguri S, Taguchi K, Kobayashi A, Ichimura Y, Sou YS, Ueno I, Sakamoto A, Tong KI *et al*: The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol* 2010, 12(3):213-223.
56. Ma J, Cai H, Wu T, Sobhian B, Huo Y, Alcivar A, Mehta M, Cheung KL, Ganesan S, Kong AN *et al*: PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function. *Mol Cell Biol* 2012, 32(8):1506-1517.
57. Rojo de la Vega M, Chapman E, Zhang DD: NRF2 and the Hallmarks of Cancer. *Cancer Cell* 2018, 34(1):21-43.
58. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018, 68(6):394-424.

59. Ahmad A: Pathways to Breast Cancer Recurrence. *ISRN Oncology* 2013, 2013:290568.
60. American Cancer Society: Cancer Facts & Figures 2021. In: *Cancer Facts & Figures.* American Cancer society, Atlanta, Georgia; 2021.
61. Neckmann U, Wolowczyk C, Aure MR, Hall M, Johannessen B, Zhao S, Skotheim RI, Andersen SB, Zwiggelaar R, Steigedal TS *et al*: A breast cancer specific NRF2 driven oxidative stress response that predicts aggressive development. In.: Norwegian University of Science and Technology; 2020.
62. Dittmer J: Mechanisms governing metastatic dormancy in breast cancer. *Semin Cancer Biol* 2017, 44:72-82.
63. Fox DB, Garcia NMG, McKinney BJ, Lupo R, Noteware LC, Newcomb R, Liu J, Locasale JW, Hirschey MD, Alvarez JV: NRF2 activation promotes the recurrence of dormant tumour cells through regulation of redox and nucleotide metabolism. *Nat Metab* 2020, 2(4):318-334.
64. Onodera Y, Motohashi H, Takagi K, Miki Y, Shibahara Y, Watanabe M, Ishida T, Hirakawa H, Sasano H, Yamamoto M *et al*: NRF2 immunolocalization in human breast cancer patients as a prognostic factor. *Endocr Relat Cancer* 2014, 21(2):241-252.
65. Kang HJ, Yi YW, Hong YB, Kim HJ, Jang YJ, Seong YS, Bae I: HER2 confers drug resistance of human breast cancer cells through activation of NRF2 by direct interaction. *Sci Rep* 2014, 4:7201.
66. Geertz M, Maerkl SJ: Experimental strategies for studying transcription factor-DNA binding specificities. *Brief Funct Genomics* 2010, 9(5-6):362-373.
67. Mundade R, Ozer HG, Wei H, Prabhu L, Lu T: Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* 2014, 13(18):2847-2852.
68. Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, Dippel D, Eesa M, Fischer U, Hausegger K, Hirsch JA *et al*: Multisociety Consensus Quality Improvement Revised Consensus Statement for Endovascular Therapy of Acute Ischemic Stroke. *Int J Stroke* 2018, 13(6):612-632.
69. Park PJ: ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* 2009, 10(10):669-680.
70. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P *et al*: ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* 2012, 22(9):1813-1831.
71. Invitrogen: A step-by-step guide to successful chromatin immunoprecipitation (ChIP) assays. In.: Thermo Fisher Scientific; 2016.
72. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P *et al*: ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* 2012, 22(9):1813-1831.
73. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A *et al*: Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 2007, 4(8):651-657.

74. Precise analysis of DNA-protein binding sequences [https://www.illumina.com/techniques/sequencing/dna-sequencing/chip-seq.html]
75. Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH: Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res* 1978, 38(10):3174-3181.
76. Miller BE, Roi LD, Howard LM, Miller FR: Quantitative selectivity of contact-mediated intercellular communication in a metastatic mouse mammary tumor line. *Cancer Res* 1983, 43(9):4102-4107.
77. Heppner GH, Dexter DL, DeNucci T, Miller FR, Calabresi P: Heterogeneity in drug sensitivity among tumor cell subpopulations of a single mammary tumor. *Cancer Res* 1978, 38(11 Pt 1):3758-3763.
78. Marhenke S, Lamlé J, Buitrago-Molina LE, Cañón JM, Geffers R, Finegold M, Sporn M, Yamamoto M, Manns MP, Grompe M *et al*: Activation of nuclear factor E2-related factor 2 in hereditary tyrosinemia type 1 and its role in survival and tumor development. *Hepatology* 2008, 48(2):487-496.
79. Satoh H, Moriguchi T, Taguchi K, Takai J, Maher JM, Suzuki T, Winnard PT, Jr., Raman V, Ebina M, Nukiwa T *et al*: Nrf2-deficiency creates a responsive microenvironment for metastasis to the lung. *Carcinogenesis* 2010, 31(10):1833-1843.
80. Chromatin Immunoprecipitation (ChIP) Assay Procedure and Essential Tools [https://www.thermofisher.com/no/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/chromatin-ip-chip-assays.html]
81. EZ-ChIP™ FAQ [https://www.merckmillipore.com/NO/en/product/EZ-ChIP,MM_NF-17-371?ReferrerURL=https%3A%2F%2Fwww.researchgate.net%2F&bd=1#anchor_FAQ]
82. 6 Tips for Immunoprecipitation [https://rockland-inc.com/immunoprecipitation-tips.aspx]
83. Normal Rabbit IgG #2729 [https://www.cellsignal.com/products/primary-antibodies/normal-rabbit-igg/2729]
84. Weinmann AS, Farnham PJ: Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 2002, 26(1):37-47.
85. Histone H3 (D2B12) XP® Rabbit mAb (ChIP Formulated) #4620 [https://www.cellsignal.com/products/primary-antibodies/histone-h3-d2b12-xp-rabbit-mab-chip-formulated/4620]
86. NRF2 (D1Z9C) XP® Rabbit mAb #12721 [https://www.cellsignal.com/products/primary-antibodies/nrf2-d1z9c-xp-rabbit-mab/12721]
87. Malhotra D, Portales-Casamar E, Singh A, Srivastava S, Arenillas D, Happel C, Shyr C, Wakabayashi N, Kensler TW, Wasserman WW *et al*: Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis. *Nucleic Acids Res* 2010, 38(17):5718-5734.

88. SimpleChIP® Mouse RPL30 Intron 2 Primers #7015 [https://www.cellsignal.com/products/chip-kits-reagents/mouse-rpl30-intron-2-primers/7015]
89. ChIP Analysis Percent Input Method [https://www.thermofisher.com/no/en/home/life-science/epigenetics-noncoding-rna-research/chromatin-remodeling/chromatin-immunoprecipitation-chip/chip-analysis.html]
90. Life Technologies: Qubit® dsDNA HS Assay Kits. In.: Thermo Fisher Scientific Inc.; 2015.
91. Qiagen: QIAseq Ultralow Input Library Kit Handbook. In. Hilden, Germany: Qiagen; 2016.
92. Agilent Technologies I: Agilent High Sensitivity DNA Kit Guide. 2013.
93. Illumina: NextSeq System Denature and Dilute Libraries Guide. In.: Illumina; 2018.
94. Illumina: NextSeq System: Denature and Dilute Libraries Guide. 2018.
95. Ewels P, Magnusson M, Lundin S, Käller M: MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016, 32(19):3047-3048.
96. Chen S, Zhou Y, Chen Y, Gu J: fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018, 34(17):i884-i890.
97. FASTQ files explained [https://support.illumina.com/bulletins/2016/04/fastq-files-explained.html]
98. Illumina: Quality Scores for Next-Generation Sequencing. 2011.
99. Illumina: Understanding Illumina Quality Scores. 2014.
100. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012, 9(4):357-359.
101. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al*: Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008, 9(9):R137.
102. Amemiya HM, Kundaje A, Boyle AP: The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* 2019, 9(1):9354.
103. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005, 102(43):15545-15550.
104. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A *et al*: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016, 44(W1):W90-97.
105. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A: Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013, 14:128.
106. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd, Barretina J, Gelfand ET, Bielski CM, Li H *et al*: Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 2019, 569(7757):503-508.

107. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D *et al*: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012, 483(7391):603-607.
108. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E *et al*: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013, 6(269):pl1.
109. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012, 2(5):401-404.
110. Kemmerer ZA, Ader NR, Mulroy SS, Eggler AL: Comparison of human Nrf2 antibodies: A tale of two proteins. *Toxicol Lett* 2015, 238(2):83-89.
111. Anti-COX IV antibody - Mitochondrial Loading Control (ab16056) [https://www.abcam.com/cox-iv-antibody-mitochondrial-loading-control-ab16056.html]
112. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A: ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 2010, 26(19):2438-2444.
113. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 2011, 9(4):e1001046.
114. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004, 306(5696):636-640.
115. Creating the gene ontology resource: design and implementation. *Genome Res* 2001, 11(8):1425-1433.
116. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015, 43(Database issue):D1049-1056.
117. Sheehan D, Meade G, Foley VM, Dowd CA: Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J* 2001, 360(Pt 1):1-16.
118. Park PJ: ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009, 10(10):669-680.
119. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, Xue P, Pi J, Kleeberger SR, Bell DA: Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha. *Nucleic Acids Res* 2012, 40(15):7416-7429.
120. He F, Antonucci L, Karin M: NRF2 as a regulator of cell metabolism and inflammation in cancer. *Carcinogenesis* 2020, 41(4):405-416.
121. Wang XJ, Sun Z, Villeneuve NF, Zhang S, Zhao F, Li Y, Chen W, Yi X, Zheng W, Wondrak GT *et al*: Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. *Carcinogenesis* 2008, 29(6):1235-1243.
122. Naegle K, Gough NR, Yaffe MB: Criteria for biological reproducibility: what does "n" mean? *Sci Signal* 2015, 8(371):fs7.
123. Zhong J, Ye Z, Lenz SW, Clark CR, Bharucha A, Farrugia G, Robertson KD, Zhang Z, Ordog T, Lee J-H: Purification of nanogram-range immunoprecipitated DNA in ChIP-

seq application. *BMC Genomics* 2017, 18(1).
124. Illumina: TruSeq™ DNA PCR-Free. 2017.
125. Wingett SW, Andrews S: FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 2018, 7:1338.
126. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 2002, 12(6):996-1006.
127. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2018, 46(D1):D8-d13.
128. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R *et al*: Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020, 583(7818):699-710.
129. Sebastián VP, Salazar GA, Coronado-Arrázola I, Schultz BM, Vallejos OP, Berkowitz L, Álvarez-Lobos MM, Riedel CA, Kalergis AM, Bueno SM: Heme Oxygenase-1 as a Modulator of Intestinal Inflammation Development and Progression. *Front Immunol* 2018, 9:1956.
130. Wilbanks EG, Facciotti MT: Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS One* 2010, 5(7):e11471.
131. Ross D, Siegel D: The diverse functionality of NQO1 and its roles in redox control. *Redox Biology* 2021, 41:101950.
132. Tonelli C, Chio IIC, Tuveson DA: Transcriptional Regulation by Nrf2. *Antioxidants & Redox Signaling* 2018, 29(17):1727-1745.
133. Yuan X, Fleming MD, Hamza I: Heme transport and erythropoiesis. *Curr Opin Chem Biol* 2013, 17(2):204-211.
134. Campbell MR, Karaca M, Adamski KN, Chorley BN, Wang X, Bell DA: Novel hematopoietic target genes in the NRF2-mediated transcriptional pathway. *Oxid Med Cell Longev* 2013, 2013:120305.
135. Shamblott MJ, Bugg EM, Lawler AM, Gearhart JD: Craniofacial abnormalities resulting from targeted disruption of the murineSim2 gene. *Developmental Dynamics* 2002, 224(4):373-380.
136. Chen D, Xu L, Li X, Chu Y, Jiang M, Xu B, Zhao M, Wang W, Wang H, Kang H *et al*: Enah overexpression is correlated with poor survival and aggressive phenotype in gastric cancer. *Cell Death & Disease* 2018, 9(10).
137. PPBP pro-platelet basic protein [Homo sapiens (human)] [https://www.ncbi.nlm.nih.gov/gene/5473]
138. CXCL3 C-X-C motif chemokine ligand 3 [Homo sapiens (human)] [https://www.ncbi.nlm.nih.gov/gene/2921]
139. Huang R, Grishagin I, Wang Y, Zhao T, Greene J, Obenauer JC, Ngan D, Nguyen D-T, Guha R, Jadhav A *et al*: The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics. *Frontiers in Pharmacology* 2019, 10.
140. Kanehisa M: Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019, 28(11):1947-1951.

141. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000, 28(1):27-30.
142. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M: KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 2021, 49(D1):D545-d551.
143. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A: The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016, 2016.
144. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Jill, Tamayo P: The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems* 2015, 1(6):417-425.
145. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S *et al*: PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 2003, 31(1):334-341.
146. Mi H, Muruganujan A, Thomas PD: PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 2013, 41(Database issue):D377-386.
147. Satoh K, Itoh K, Yamamoto M, Tanaka M, Hayakari M, Ookawa K, Yamazaki T, Sato T, Tsuchida S, Hatayama I: Nrf2 transactivator-independent GSTP1-1 expression in "GSTP1-1 positive" single cells inducible in female mouse liver by DEN: a preneoplastic character of possible initiated cells. *Carcinogenesis* 2002, 23(3):457-462.
148. Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, Massard C, Lévy C, Arnedos M, Lacroix-Triki M *et al*: Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med* 2016, 13(12):e1002201.
149. Ciriello G, Michael, Andrew, Matthew, Suhn, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C *et al*: Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 2015, 163(2):506-519.
150. Dixon RAF, Diehl RE, Opas E, Rands E, Vickers PJ, Evans JF, Gillard JW, Miller DK: Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. *Nature* 1990, 343(6255):282-284.
151. He F, Ru X, Wen T: NRF2, a Transcription Factor for Stress Response and Beyond. *International Journal of Molecular Sciences* 2020, 21(13):4777.
152. Hu Q, Ren J, Li G, Wu J, Wu X, Wang G, Gu G, Ren H, Hong Z, Li J: The mitochondrially targeted antioxidant MitoQ protects the intestinal barrier by ameliorating mitochondrial DNA damage via the Nrf2/ARE signaling pathway. *Cell Death & Disease* 2018, 9(3).
153. Lim PJ, Duarte TL, Arezes J, Garcia-Santos D, Hamdi A, Pasricha SR, Armitage AE, Mehta H, Wideman S, Santos AG *et al*: Nrf2 controls iron homeostasis in haemochromatosis and thalassaemia via Bmp6 and hepcidin. *Nat Metab* 2019, 1(5):519-531.
154. Zhou X-L, Zhu C-Y, Wu Z-G, Guo X, Zou W: The oncoprotein HBXIP competitively binds KEAP1 to activate NRF2 and enhance breast cancer cell growth and metastasis. *Oncogene* 2019, 38(21):4028-4046.

155. Onodera Y, Motohashi H, Takagi K, Miki Y, Shibahara Y, Watanabe M, Ishida T, Hirakawa H, Sasano H, Yamamoto M *et al*: NRF2 immunolocalization in human breast cancer patients as a prognostic factor. *Endocrine-Related Cancer* 2014, 21(2):241-252.
156. Giudice A, Barbieri A, Bimonte S, Cascella M, Cuomo A, Crispo A, D'Arena G, Galdiero M, Della Pepa ME, Botti G *et al*: Dissecting the prevention of estrogen-dependent breast carcinogenesis through Nrf2-dependent and independent mechanisms. *Onco Targets Ther* 2019, 12:4937-4953.
157. Wu S, Lu H, Bai Y: Nrf2 in cancers: A double-edged sword. *Cancer Med* 2019, 8(5):2252-2267.
158. Ma Q: Role of nrf2 in oxidative stress and toxicity. *Annu Rev Pharmacol Toxicol* 2013, 53:401-426.
159. Taxak N, Bharatam PV: Drug metabolism. *Resonance* 2014, 19(3):259-282.
160. auf dem Keller U, Kümin A, Braun S, Werner S: Reactive oxygen species and their detoxification in healing skin wounds. *J Investig Dermatol Symp Proc* 2006, 11(1):106-111.
161. Leier I, Jedlitschky G, Buchholz U, Cole SP, Deeley RG, Keppler D: The MRP gene encodes an ATP-dependent export pump for leukotriene C4 and structurally related conjugates. *J Biol Chem* 1994, 269(45):27807-27810.
162. Taguchi K, Yamamoto M: The KEAP1–NRF2 System as a Molecular Target of Cancer Treatment. *Cancers* 2020, 13(1):46.
163. Carlisi D, De Blasio A, Drago-Ferrante R, Di Fiore R, Buttitta G, Morreale M, Scerri C, Vento R, Tesoriere G: Parthenolide prevents resistance of MDA-MB231 cells to doxorubicin and mitoxantrone: the role of Nrf2. *Cell Death Discov* 2017, 3:17078.
164. Luo Y, Zhang W, Xu L, Chen Y, Xu Y, Yuan L: Long Non-Coding RNA PVT1 Regulates the Resistance of the Breast Cancer Cell Line MDA-MB-231 to Doxorubicin via Nrf2. *Technol Cancer Res Treat* 2020, 19:1533033820980763.
165. Kageyama S, Hanada E, Ii H, Tomita K, Yoshiki T, Kawauchi A: Gamma-Glutamylcyclotransferase: A Novel Target Molecule for Cancer Diagnosis and Treatment. *BioMed Research International* 2015, 2015:1-5.
166. Matsumura K, Nakata S, Taniguchi K, Ii H, Ashihara E, Kageyama S, Kawauchi A, Yoshiki T: Depletion of γ-glutamylcyclotransferase inhibits breast cancer cell growth via cellular senescence induction mediated by CDK inhibitor upregulation. *BMC Cancer* 2016, 16(1):748.
167. Ii H, Yoshiya T, Nakata S, Taniguchi K, Hidaka K, Tsuda S, Mochizuki M, Nishiuchi Y, Tsuda Y, Ito K *et al*: A Novel Prodrug of a γ-Glutamylcyclotransferase Inhibitor Suppresses Cancer Cell Proliferation in vitro and Inhibits Tumor Growth in a Xenograft Mouse Model of Prostate Cancer. *ChemMedChem* 2018, 13(2):155-163.
168. Kageyama S, Isono T, Iwaki H, Wakabayashi Y, Okada Y, Kontani K, Yoshimura K, Terai A, Arai Y, Yoshiki T: Identification by proteomic analysis of calreticulin as a marker for bladder cancer and evaluation of the diagnostic accuracy of its detection in urine. *Clin Chem* 2004, 50(5):857-866.
169. Huang Q, Zhou Y, Li Y, Liao Z: GGCT promotes colorectal cancer migration and

invasion via epithelial-mesenchymal transition. *Oncology Letters* 2020, 20(2):1063-1070.
170. Whiteside TL: The tumor microenvironment and its role in promoting tumor growth. *Oncogene* 2008, 27(45):5904-5912.
171. Coussens LM, Werb Z: Inflammation and cancer. *Nature* 2002, 420(6917):860-867.
172. CSF1 colony stimulating factor 1 [Homo sapiens (human)] [https://www.ncbi.nlm.nih.gov/gene/]
173. RICHARDSEN E, UGLEHUS RD, JOHNSEN SH, BUSUND L-T: Macrophage-Colony Stimulating Factor (CSF1) Predicts Breast Cancer Progression and Mortality. *Anticancer Research* 2015, 35(2):865-874.
174. Cassetta L, Fragkogianni S, Sims AH, Swierczak A, Forrester LM, Zhang H, Soong DYH, Cotechini T, Anur P, Lin EY *et al*: Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. *Cancer Cell* 2019, 35(4):588-602.e510.
175. Smith NLD, Bromley MJ, Denning DW, Simpson A, Bowyer P: Elevated Levels of the Neutrophil Chemoattractant Pro–Platelet Basic Protein in Macrophages From Individuals With Chronic and Allergic Aspergillosis. *The Journal of Infectious Diseases* 2015, 211(4):651-660.
176. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A: Neutrophil Function: From Mechanisms to Disease. *Annual Review of Immunology* 2012, 30(1):459-489.
177. Li L, Zhang L, Tian Y, Zhang T, Duan G, Liu Y, Yin Y, Hua D, Qi X, Mao Y: Serum Chemokine CXCL7 as a Diagnostic Biomarker for Colorectal Cancer. *Front Oncol* 2019, 9:921.
178. Guo Q, Jian Z, Jia B, Chang L: CXCL7 promotes proliferation and invasion of cholangiocarcinoma cells. *Oncology Reports* 2017, 37(2):1114-1122.
179. Bièche I, Chavey C, Andrieu C, Busson M, Vacher S, Le Corre L, Guinebretière JM, Burlinchon S, Lidereau R, Lazennec G: CXC chemokines located in the 4q21 region are up-regulated in breast cancer. *Endocr Relat Cancer* 2007, 14(4):1039-1052.
180. Wolowczyk CI, Neckmann U, Sofias AM, Giambelluca M, Ostrop J, Mildenberger JM, Prestvik WS, Nonstad U, Hak S, Bjørkøy G: Comparing breast cancer cells of the 4T1 model identifies CXCL3 secretion to recruit neutrophil-like cells and establish a pro-metastatic microenvironment. In. Trondheim, Norway: Norwegian University of Science and Technology; 2020.
181. Neuhofer W: Role of NFAT5 in inflammatory disorders associated with osmotic stress. *Curr Genomics* 2010, 11(8):584-590.
182. López-Rodríguez C, Aramburu J, Jin L, Rakeman AS, Michino M, Rao A: Bridging the NFAT and NF-kappaB families: NFAT5 dimerization regulates cytokine gene transcription in response to osmotic stress. *Immunity* 2001, 15(1):47-58.
183. Cruceriu D, Baldasici O, Balacescu O, Berindan-Neagoe I: The dual role of tumor necrosis factor-alpha (TNF-α) in breast cancer: molecular insights and therapeutic approaches. *Cellular Oncology* 2020, 43(1):1-18.
184. Fougère M, Gaudineau B, Barbier J, Guaddachi F, Feugeas JP, Auboeuf D, Jauliac S:

NFAT3 transcription factor inhibits breast cancer cell motility by targeting the Lipocalin 2 gene. *Oncogene* 2010, 29(15):2292-2301.
185. Remo A, Simeone I, Pancione M, Parcesepe P, Finetti P, Cerulo L, Bensmail H, Birnbaum D, Van Laere SJ, Colantuoni V *et al*: Systems biology analysis reveals NFAT5 as a novel biomarker and master regulator of inflammatory breast cancer. *Journal of Translational Medicine* 2015, 13(1).
186. Tegowski M, Baldwin A: Noncanonical NF-κB in Cancer. *Biomedicines* 2018, 6(2):66.
187. Mantovani A, Barajon I, Garlanda C: IL-1 and IL-1 regulatory pathways in cancer progression and therapy. *Immunological Reviews* 2018, 281(1):57-61.
188. Zhang W, Borcherding N, Kolb R: IL-1 Signaling in Tumor Microenvironment. In.: Springer International Publishing; 2020: 1-23.
189. Plitas G, Konopacki C, Wu K, Bos PD, Morrow M, Ekaterina, Dmitriy, Alexander: Regulatory T Cells Exhibit Distinct Features in Human Breast Cancer. *Immunity* 2016, 45(5):1122-1134.
190. De Simone M, Arrigoni A, Rossetti G, Gruarin P, Ranzani V, Politano C, Bonnal RJP, Provasi E, Sarnicola ML, Panzeri I *et al*: Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. *Immunity* 2016, 45(5):1135-1147.
191. Bos PD, Plitas G, Rudra D, Lee SY, Rudensky AY: Transient regulatory T cell ablation deters oncogene-driven breast cancer and enhances radiotherapy. *J Exp Med* 2013, 210(11):2435-2466.
192. Korolnek T, Hamza I: Macrophages and iron trafficking at the birth and death of red cells. *Blood* 2015, 125(19):2893-2897.
193. Kumar S, Bandyopadhyay U: Free heme toxicity and its detoxification systems in human. *Toxicol Lett* 2005, 157(3):175-188.
194. Simmons WR, Wain L, Toker J, Jagadeesh J, Garrett LJ, Pek RH, Hamza I, Bodine DM: Normal Iron Homeostasis Requires the Transporter SLC48A1 for Efficient Heme-Iron Recycling in Mammals. *Frontiers in Genome Editing* 2020, 2.
195. Gbotosho OT, Ghosh S, Kapetanaki MG, Lin Y, Weidert F, Bullock GC, Ofori-Acquah SF, Kato GJ: Cardiac expression of HMOX1 and PGF in sickle cell mice and haem-treated wild type mice dominates organ expression profiles via Nrf2 (Nfe2l2). *Br J Haematol* 2019, 187(5):666-675.
196. Fogarty FM, O'Keeffe J, Zhadanov A, Papkovsky D, Ayllon V, O'Connor R: HRG-1 enhances cancer cell invasive potential and couples glucose metabolism to cytosolic/extracellular pH gradient regulation by the vacuolar-H+ ATPase. *Oncogene* 2014, 33(38):4653-4663.
197. Transcription Factors. In: *Reference Module in Biomedical Sciences*. Elsevier; 2014.
198. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R *et al*: Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020, 583(7818):699-710.
199. Rose AB: Introns as Gene Regulators: A Brick on the Accelerator. *Front Genet* 2018, 9:672.

200. Canet MJ, Merrell MD, Harder BG, Maher JM, Wu T, Lickteig AJ, Jackson JP, Zhang DD, Yamamoto M, Cherrington NJ: Identification of a functional antioxidant response element within the eighth intron of the human ABCC3 gene. *Drug Metab Dispos* 2015, 43(1):93-99.
201. Arya M, Shergill IS, Williamson M, Gommersall L, Arya N, Patel HR: Basic principles of real-time quantitative PCR. *Expert Rev Mol Diagn* 2005, 5(2):209-219.
202. NextSeq 550 System Frequently Asked Questions [https://support.illumina.com/sequencing/sequencing_instruments/nextseq-550/questions.html]
203. How short inserts affect sequencing performance [https://support.illumina.com/bulletins/2020/12/how-short-inserts-affect-sequencing-performance.html]
204. Tan G, Opitz L, Schlapbach R, Rehrauer H: Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports* 2019, 9(1):2856.
205. Kidder BL, Hu G, Zhao K: ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol* 2011, 12(10):918-922.
206. Diagenode: Guide for successful chromatin preparation using the Bioruptor® Pico: Sonication. 2015.
207. Fragmentation of chromatin in ChIP [https://www.sigmaaldrich.com/technical-documents/articles/biology/fragmentation-of-chromatin-in-chip.html]
208. Skene PJ, Henikoff S: A simple method for generating high-resolution maps of genome-wide protein binding. *eLife* 2015, 4.
209. Chorley BN, Campbell MR, Wang X, Karaca M, Sambandan D, Bangura F, Xue P, Pi J, Kleeberger SR, Bell DA: Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha. *Nucleic Acids Research* 2012, 40(15):7416-7429.
210. Yu C-P, Lin J-J, Li W-H: Positional distribution of transcription factor binding sites in Arabidopsis thaliana. *Scientific Reports* 2016, 6(1):25164.
211. Becks L, Prince M, Burson H, Christophe C, Broadway M, Itoh K, Yamamoto M, Mathis M, Orchard E, Shi R *et al*: Aggressive mammary carcinoma progression in Nrf2 knockout mice treated with 7,12-dimethylbenz[a]anthracene. *BMC Cancer* 2010, 10(1):540.
212. Surh Y, Kundu J, Na H: Nrf2 as a master redox switch in turning on the cellular signaling involved in the induction of cytoprotective genes by some chemopreventive phytochemicals. *Planta Medica* 2008, 74:1526-1539.
213. Deighton RF, Markus NM, Al-Mubarak B, Bell KFS, Papadia S, Meakin PJ, Chowdhry S, Hayes JD, Hardingham GE: Nrf2 target genes can be controlled by neuronal activity in the absence of Nrf2 and astrocytes. *Proceedings of the National Academy of Sciences* 2014, 111(18):E1818-E1820.
214. Huppke P, Weissbach S, Church JA, Schnur R, Krusen M, Dreha-Kulaczewski S, Kühn-Velten WN, Wolf A, Huppke B, Millan F *et al*: Activating de novo mutations in NFE2L2 encoding NRF2 cause a multisystem disorder. *Nature Communications* 2017,

8(1).
215. Xu J, Kudron MM, Victorsen A, Gao J, Ammouri HN, Navarro FCP, Gevirtzman L, Waterston RH, White KP, Reinke V *et al*: To mock or not: a comprehensive comparison of mock IP and DNA input for ChIP-seq. In.: Cold Spring Harbor Laboratory; 2019.
216. NRF2 (D1Z9C) XP® Rabbit mAb #12721 [https://www.cellsignal.com/products/primary-antibodies/nrf2-d1z9c-xp-rabbit-mab/12721]
217. Lau A, Tian W, Whitman SA, Zhang DD: The predicted molecular weight of Nrf2: it is what it is not. *Antioxid Redox Signal* 2013, 18(1):91-93.
218. Zhang HS, Zhang ZG, Du GY, Sun HL, Liu HY, Zhou Z, Gou XM, Wu XH, Yu XY, Huang YH: Nrf2 promotes breast cancer cell migration via up-regulation of G6PD/HIF-1α/Notch1 axis. *J Cell Mol Med* 2019, 23(5):3451-3463.
219. Zhou Z, Xu J, Bao X, Shi J, Liu B, Chen Y, Li J: Nuclear Nrf2 Activity in Laryngeal Carcinoma is Regulated by SENP3 After Cisplatin-Induced Reactive Oxygen Species Stress. *J Cancer* 2019, 10(15):3427-3434.

7 Appendices

7.1 Supplementary Figures and Tables

List of Supplementary Figures

List of Supplementary Tables

Supplementary Table 1. Primer sequences for ChIP-qPCR to evaluate quality of ChIP samples prior to sequencing. The sequences were referenced from a NRF2 ChIP-seq study by Malhotra et al. *Txnrd1-P1 and Txnrd1-P2 represent two different NRF2-binding sites within the gene* Txnrd1 [87].

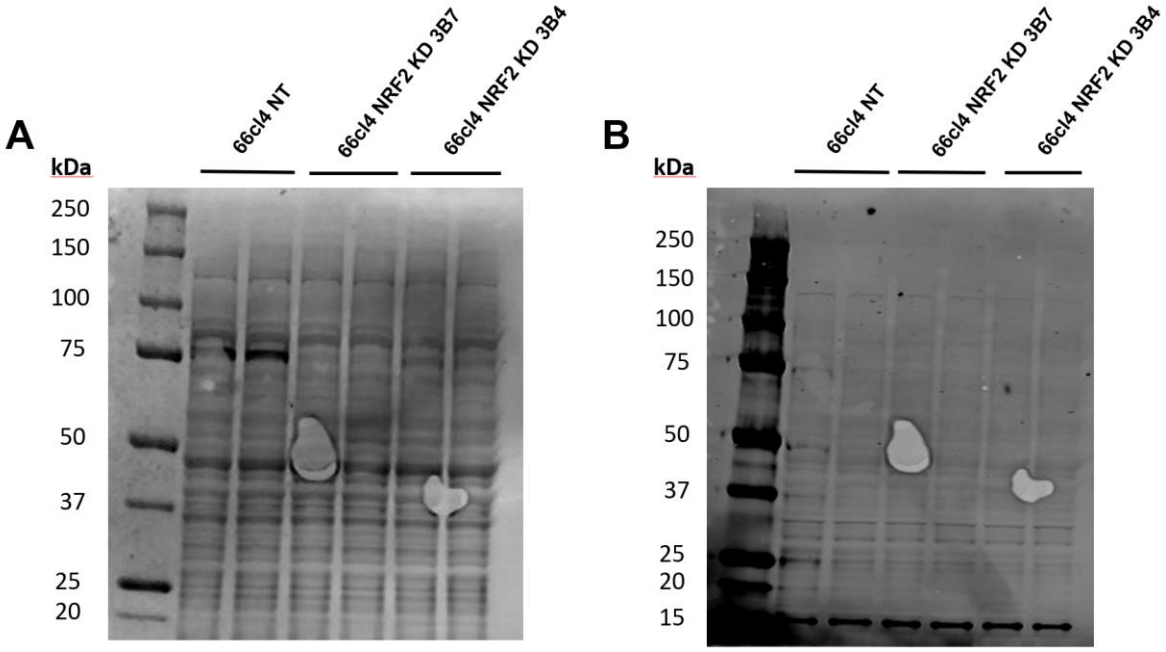
Gene	
Nqo1	S-5'-AGTCACCTTTGCACGCTAGG-3'
	AS-5'-TCTAAGAGCAGAACGCAGCA-3'
Txnrd1_P1	S-5'-CAGCGAGGATGTTGTACGG-3'
	AS-5'-TAAAGAGCTGCGGGTTCCTA-3'
Txnrd1-P2	S-5'-ACAGTGTCGCCAAGCATTTA-3'
	AS-5'-ACGCCTGCTCAGTGCTAAGT-3'
Gsta3	S-5'-ATCCTCATTGACAGCCGAAG-3'
	AS-5'-CAGTGTTCAGAGGCCAGACA-3'
Gsta4	S-5'-GGTTCCCCCGTAGAGAATGT-3'
	AS-5'-CACGTTGGTTAACGCTTTGTT-3'

S = sense, AS = antisense

Supplementary Table 2. Primer sequences for ChIP-qPCR to verify ChIP-seq data. The sequences were extracted from nine significant ChIP-seq peak regions and tailored using an online PCR primer design tool called GenScript (https://www.genscript.com/tools/pcr-primers-designer).

Gene	
Slc48a1	S-5'-TGACTCAGCATTTTGCAGCCCA-3'
	AS-5'-CAGCATGCCCGACAGCCACA-3'
Hmox1	S-5'-CGCAGCGGCTGGAATGCTGA-3'
	AS-5'-CCCCTCCCAACCATGACACAGC-3'
Fth1	S-5'-CCACAGCCCTCCAGGCCTTC-3'
	AS-5'-GCTCCTTTGGAGGGGTTGGGC-3'
Slc40a1	S-5'-TGTTACTGTGAGGCTTGGGTTGTTCT-3'
	AS-5'-ACACATTTCAGCTGGGGAAAATGCAA-3'
Rpa1	S-5'-CCCATGTCCTCTGGAAGAACAGCC-3'
	AS-5'-CCAGCTTGGTGGGGCTGGAG-3'
Abcc1	S-5'-CCCTGGCTTGAGTTAGGGCGA-3'
	AS-5'-GCCACAGTTCCTCCCAGACCG-3'
Krt9	S-5'-CCGCCTGAGCTTTCTGGCCC-3'
	AS-5'-GCCAGCATCCAGCCTCACGA-3'
Sema3e	S-5'-ACGGCATTGACAAGACAGTCGG-3'
	AS-5'-ATGCACCCTGTCCTGCCCCT-3'
Nbea	S-5'-TCTGAGCCAGCCACTTCAGTCA-3'
	AS-5'-TGCCCGGTGGGGCTGTTGAT-3'
Tkt	S-5'-AGCCCCAAGCAGGAAGCATGA-3'
	AS-5'-CCCAGGGCCAGGCAGTCAGA-3'

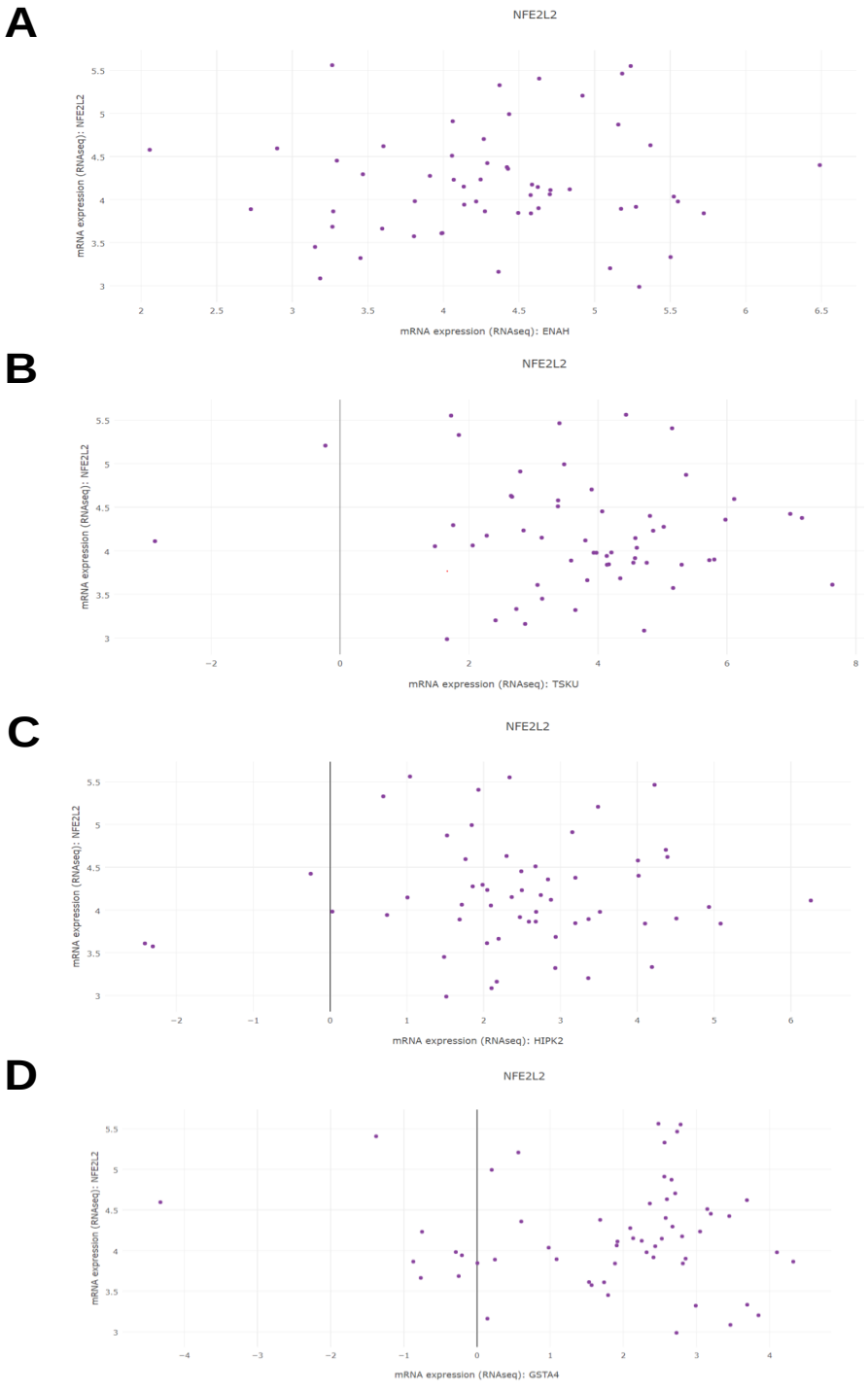
S = sense, AS = antisense



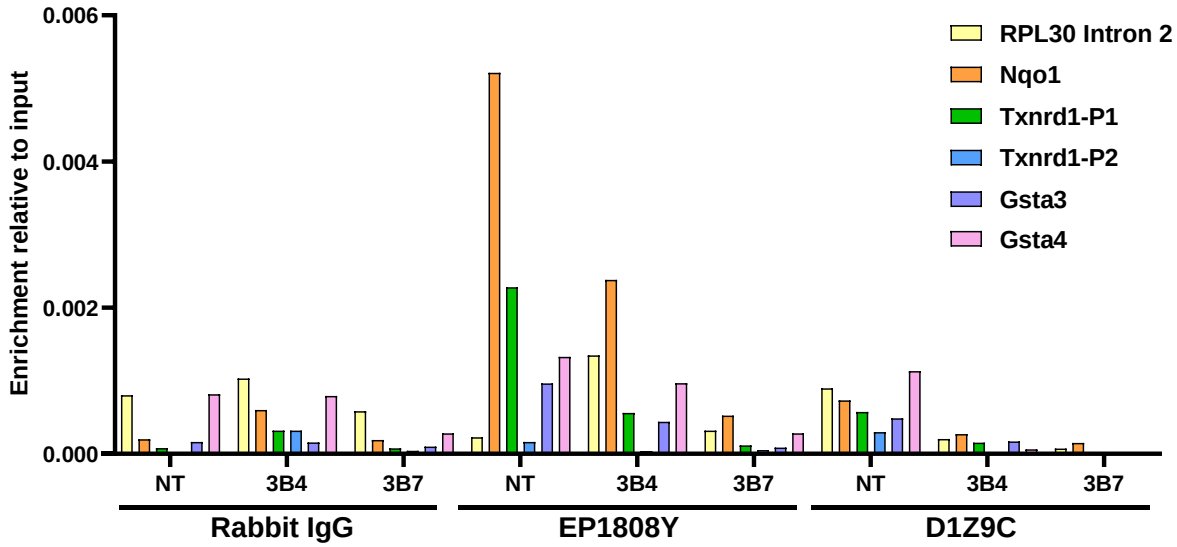
Supplementary Figure 1. Immunoblotted nitrocellulose membranes of 66cl4 cells. (A) NRF2 (D1Z9C) rabbit monoclonal antibody from Signaling technology was applied to the membrane and shows two strong bands sized 75 kDa in the 66cl4 NT lanes. The first band is partially weaker which was due to insufficient attachment between the gel and the membrane during dry blotting. **(B)** The same membrane was subsequently probed with anti-COX IV antibody from Abcam as a loading control and strong bands of 15 kDa are exhibited in all lanes indicating equal protein loading. The three cell line samples were each presented in two lanes as technical replicates.

Supplementary Table 3. RNA-seq results of commonly downregulated genes between NRF2 KD cell lines (66cl4 NRF2 KD 3B4, 66cl4 NRF2 KD 3B7) when compared with 66cl4 NT cell line. 40 commonly downregulated genes with the lowest adjusted p-value (padj) and low $log_2(fold\ change)$ were selected and sorted from smallest to largest in padj.

	66cl4 NRF2 KD 3B4 / 66cl4 NT		66cl4 NRF2 KD 3B7 / 66cl4 NT		Average	
Gene	$log_2(fold\ change$	**padj**	$log_2(fold\ change$	**padj**	$log_2(fold\ change$	**padj**
Nfe2l2	-2.881	1.80684E-54	-3.263	5.57476E-70	-3.072	9.03418E-55
Slc48a1	-1.324	4.64126E-49	-1.476	5.77967E-61	-1.4	2.32063E-49
Gm2a	-3.363	6.52239E-43	-3.379	6.22457E-43	-3.371	6.37348E-43
Hvcn1	-4.466	3.08882E-21	-5.806	1.9462E-26	-5.136	1.54442E-21
Serpina1d	-26.037	8.18321E-21	-25.949	1.2037E-20	-25.993	1.01101E-20
Gstk1	-4.488	1.13618E-20	-5.524	5.09819E-20	-5.006	3.11719E-20
Alox5ap	-3.323	8.03591E-15	-3.28	2.20362E-14	-3.3015	1.50361E-14
Ppp1r13b	-4.067	1.00425E-13	-4.381	3.27055E-15	-4.224	5.18479E-14
Pla2g5	-4.438	1.58915E-12	-5.405	2.25451E-18	-4.9215	7.94574E-13
Ndufa8	-1.154	2.81428E-15	-1.049	1.79811E-12	-1.1015	9.00462E-13
Enah	-1.02	4.19297E-16	-0.896	2.75905E-12	-0.958	1.37973E-12
Aldh3a2	-1.133	1.03485E-14	-1.021	8.85119E-12	-1.077	4.43077E-12
Smpdl3b	-3.566	1.20973E-10	-5.607	5.74318E-22	-4.5865	6.04865E-11
Sim2	-3.7	1.88569E-10	-4.999	2.71671E-12	-4.3495	9.56426E-11
Cldn6	-2.444	1.88569E-10	-2.445	2.39858E-10	-2.4445	2.14213E-10
Sbsn	-2.969	3.64172E-16	-2.053	1.38228E-09	-2.511	6.9114E-10
Fstl1	-1.841	1.51242E-09	-2.135	1.5637E-12	-1.988	7.56991E-10
Zfp652	-1.671	1.97028E-11	-1.526	1.68804E-09	-1.5985	8.53872E-10
Gca	-4.257	1.81379E-09	-4.608	9.58987E-11	-4.4325	9.54843E-10
Hipk2	-2.271	3.44478E-12	-1.997	2.42218E-09	-2.134	1.21281E-09
Cldn9	-1.831	2.24506E-10	-1.705	6.3099E-09	-1.768	3.2672E-09
Ugt1a6a	-4.6	3.04558E-11	-7.47	1.11758E-08	-6.035	5.60313E-09
Reck	-2.833	9.56809E-10	-2.64	1.95964E-08	-2.7365	1.02766E-08
Car9	-3.15	2.19353E-08	-4.534	1.65614E-17	-3.842	1.09677E-08
Pdgfrl	-6.486	2.19353E-08	-9.18	2.63414E-09	-7.833	1.22847E-08
Bc1	-1.953	2.7802E-08	-2.276	1.68571E-10	-2.1145	1.39853E-08
Evc2	-3.201	7.60408E-09	-5.163	2.44995E-08	-4.182	1.60518E-08
Csnk2a1	-0.586	1.58684E-13	-0.455	4.58398E-08	-0.5205	2.292E-08
Loxl1	-2.23	2.19353E-08	-2.215	3.29759E-08	-2.2225	2.74556E-08
Lrch2	-3.716	8.81502E-13	-2.88	5.51292E-08	-3.298	2.75651E-08
Mex3a	-1.406	1.25523E-10	-1.21	5.7751E-08	-1.308	2.89383E-08
Serpina1a	-12.998	7.00878E-09	-12.216	6.68179E-08	-12.607	3.69133E-08
Inpp5j	-2.535	1.16084E-07	-2.979	1.89616E-09	-2.757	5.899E-08
Cacng7	-1.907	1.07516E-07	-2.003	2.44995E-08	-1.955	6.60078E-08
Ppbp	-2.443	1.27562E-09	-1.993	1.65082E-07	-2.218	8.31791E-08
Tshz2	-4.505	2.14798E-07	-7.623	6.68179E-08	-6.064	1.40808E-07
Gsta4	-1.964	3.08308E-07	-2.282	1.42836E-09	-2.123	1.54868E-07
Tsku	-1.67	3.65053E-08	-1.575	3.0099E-07	-1.6225	1.68747E-07
Evc	-4.978	4.10625E-08	-5.946	3.02787E-07	-5.462	1.71925E-07
Csf2rb	-5.781	2.07811E-08	-6.729	5.64475E-07	-6.255	2.92628E-07



Supplementary Figure 2. Scatter plots presenting interrelation of mRNA expression between *NFE2L2* and NRF2-targeted genes in 60 breast cancer cell lines. mRNA expression levels of the seven known NRF2-regulated genes from RNA-seq were searched in Cancer Cell Line Encyclopedia (CCLE) to examine its association with *NFE2L2* expression in breast cancer [107]. Using data from 60 breast cancer cell lines in CCLE, genes **(A)** *ENAH*, **(B)** *TSKU*, **(C)** *HIPK2* and **(D)** *GSTA4* were found to have a correlation in mRNA expression with *NFE2l2*, verifying the downregulated NRF2 expression in the 66cl4 NRF2 KD cell lines when compared to 66cl4 NT cells [106, 107]. The *x* axis and *y* axis represent mRNA expression (RNA-seq) of each NRF2-regulated gene and mRNA expression (RNA-seq) of *NFE2L2*, respectively.

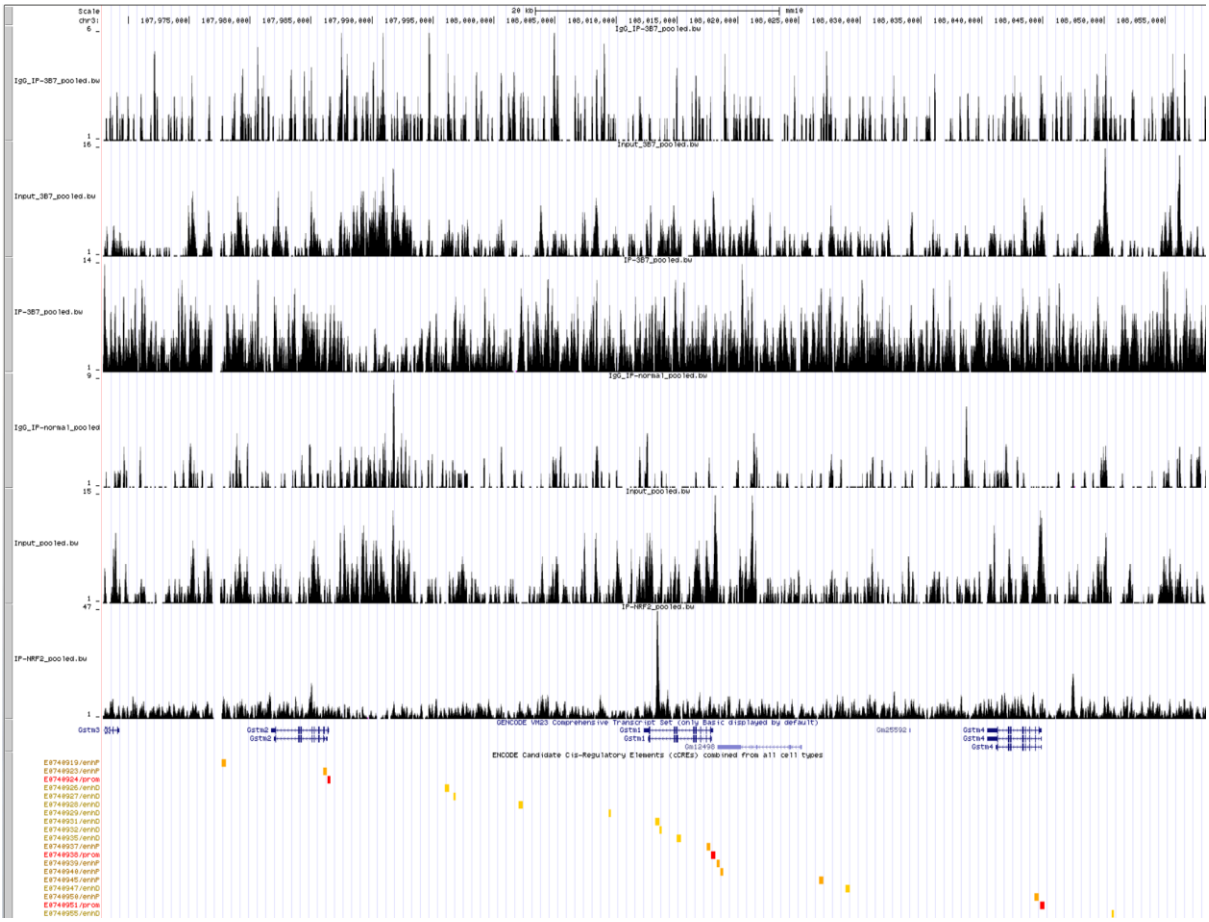


Supplementary Figure 3. *Nqo1* and *Txnrd1-P1* was chosen as the optimal primer sets for quality control of ChIP samples. Genes *Nqo1* and *Txnrd1-P1* were highly amplified in 66cl4 NT cells that was immunoprecipitated by anti-NRF2 antibody (EP1808Y) (Abcam). However, primers Txnrd1-P2, Gsta3 and Gsta4 did not exhibit significant enrichment signals. Thus, primers Nqo1 and Txnrd1-P1 were chosen to be used for assessing qualities of ChIP DNA. The primers of five NRF2-regulated genes were referenced from an NRF2 ChIP-seq study by Malhotra et al. Txnrd1-P1 and Txnrd1-P2 represent two different NRF2-binding sites within the *Txnrd1* gene [87]. Enriched DNA levels were measured using percent input method and normalized by C_T values of input samples [89]. The graph presents average values from the first and second set of ChIP-qPCR. Abbreviation: 66cl4 NT (NT), 66cl4 NRF2 KD 3B4 (3B4), 66cl4 NRF2 KD 3B7 (3B7), Normal Rabbit IgG from Cell Signaling Technology (Rabbit IgG) and NRF2 (D1Z9C) rabbit mAb from Cell Signaling Technology (D1Z9C).

Supplementary Table 4. DNA concentration of six biological ChIP replicates measured prior to library prep. The first biological ChIP replicate exhibited the lowest dsDNA concentrations and was thus excluded to be used for library prep. In the first replicate, dsDNA concentrations of the NT (66cl4 NT) and 3B7 (NRF2 KD 3B7) cell lines that were precipitated by anti-NRF2 antibody (EP1808Y) (Abcam) and normal rabbit IgG (Cell Signaling Technology) could not be detected. A two-point standard curve calculation method was applied but could not deduce an estimated concentration value likewise. To prevent low concentration in later biological replicates, samples were not diluted since the 3rd biological replicate. The concentrations were measured with Qubit™ dsDNA HS Assay Kit and Qubit® 3.0 Fluorometer at the NTNU Genomics Core Facility [90]. Symbol ○ indicates that the samples were diluted 1:2 in TE-buffer and x indicates that the samples were not diluted. Abbreviation: 66cl4 NT (NT), NRF2 KD 3B7 (3B7), Normal Rabbit IgG from Cell Signaling Technology (Rabbit IgG), immunoprecipitation by anti-NRF2 antibody (EP1808Y) from Abcam (NRF2 IP).

Biological replicate	Cell line	Sample	Dilution (1:2 in TE buffer)	Elution volume (µl)	Total volume (µl)	Concentration (ng/µl)	Estimated concentration using two-point standard curve (ng/µl)	Total amount of dsDNA (ng)
1st ChIP	NT	Input	○	50	30	20.8		624
	NT	NRF2 IP	○	50	43	low	not possible	
	NT	Normal Rabbit IgG	○	50	47	low	not possible	
	3B7	Input	○	50	31	22.4		694.4
	3B7	NRF2 IP	○	50	39	low	not possible	
	3B7	Normal Rabbit IgG	○	50	39	low	not possible	
2nd ChIP	NT	Input	○	50	42	16.2		680.4
	NT	NRF2 IP	○	50	55	low	not possible	
	NT	Normal Rabbit IgG	○	50	55	0.363		19.965
	3B7	Input	○	50	20	32		640
	3B7	NRF2 IP	○	50	56	low	not possible	
	3B7	Normal Rabbit IgG	○	50	56	low	not possible	
3rd ChIP	NT	Input	x	50	11	43.8		481.8
	NT	NRF2 IP	x	50	11	0.68		7.48
	NT	Normal Rabbit IgG	x	50	18	0.147		2.646
	3B7	Input	x	50	13	48		624
	3B7	NRF2 IP	x	50	11	low	0.04	0.44
	3B7	Normal Rabbit IgG	x	50	17.5	low	0.01	0.175
4th ChIP	NT	Input	x	30	20	27.3		546
	NT	NRF2 IP	x	30	10	3.55		35.5
	NT	Normal Rabbit IgG	x	30	12	low	0.034	0.408
	3B7	Input	x	30	18	12.8		230.4
	3B7	NRF2 IP	x	30	12	1.64		19.68
	3B7	Normal Rabbit IgG	x	30	12	low	0.016	0.192

5th ChIP	NT	Input	x	30	18	41.2		741.6
	NT	NRF2 IP	x	30	13	0.183		2.379
	NT	Normal Rabbit IgG	x	30	14	low	0.042	0.588
	3B7	Input	x	30	19	48.4		919.6
	3B7	NRF2 IP	x	30	13	0.128		1.664
	3B7	Normal Rabbit IgG	x	30	13	0.051		0.663
6th ChIP	NT	Input	x	30	12	21.7		260.4
	NT	NRF2 IP	x	30	12	0.132		1.584
	NT	Normal Rabbit IgG	x	30	12	0.051		0.612
	3B7	Input	x	30	18	39.5		711
	3B7	NRF2 IP	x	30	12	0.0334		0.4008
	3B7	Normal Rabbit IgG	x	30	12	0.054		0.648



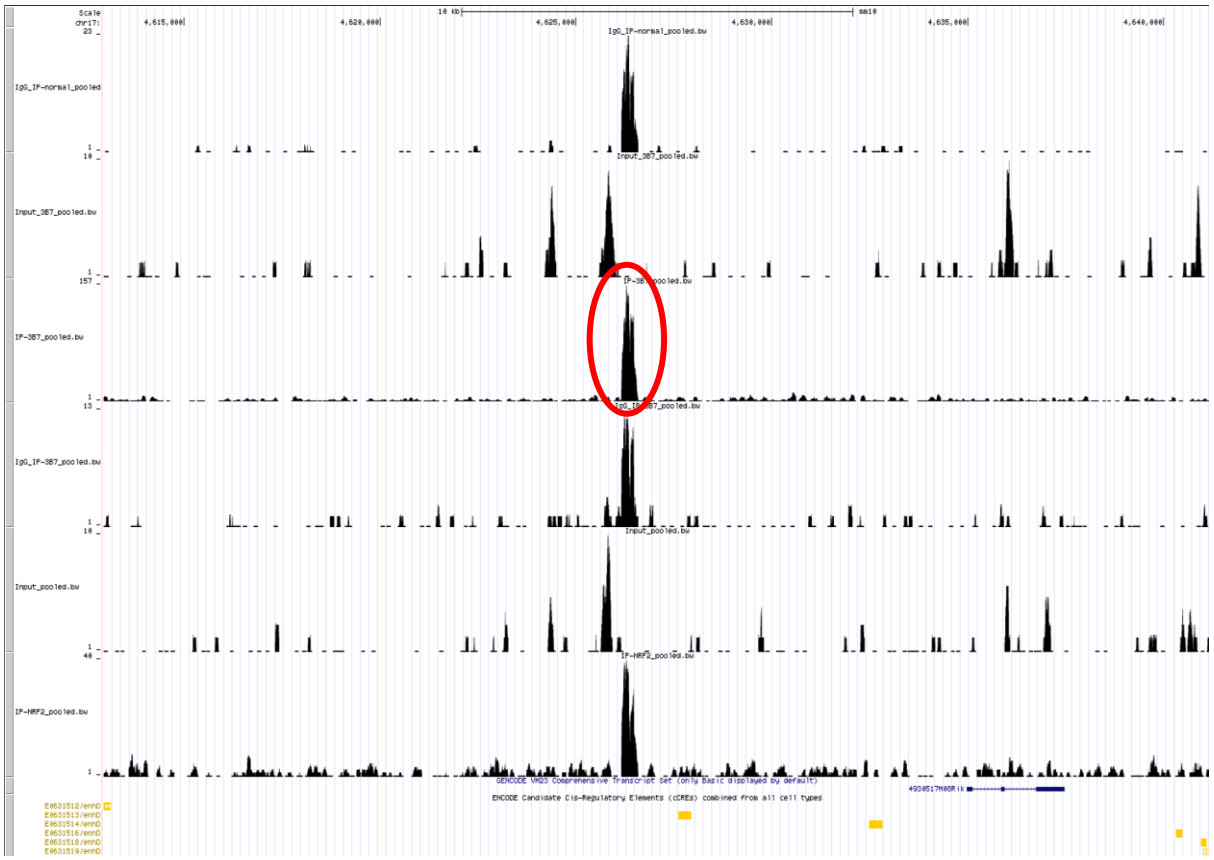
Supplementary Figure 4. Clustered ChIP-seq peaks of the chromosome 3 (108013341-108013442) region from all six types of sequenced samples are displayed. Each lane represents a sample group pooled from five biological replicates for bioinformatical analyses. From the top, the samples are the following in the consecutive order: Normal rabbit IgG-precipitated 66cl4 NRF2 KD 3B7 (3B7), input of 3B7, anti-NRF2 antibody EP1808Y (EP1808Y)-precipitated 3B7, normal rabbit IgG-precipitated 66cl4 NT (NT), input of NT, EP1808Y-precipiated NT. The EP1808Y-enriched NT group (last lane) exhibits a 47-sized peak whereas the remaining five sample groups show ground-level peak heights. The respective samples exhibited a peak height of 6, 16, 14, 9, 15 and 47. The ChIP-seq peaks were displayed with UCSC genome browser [126].

Supplementary Table 5. A total of 94 ChIP-seq peak regions identified through bioinformatic analyses of sequenced 66cl4 NT samples which were precipitated by anti-NRF2 antibody EP1808Y (Abcam). Abbreviation: Intergenic region (IGR)

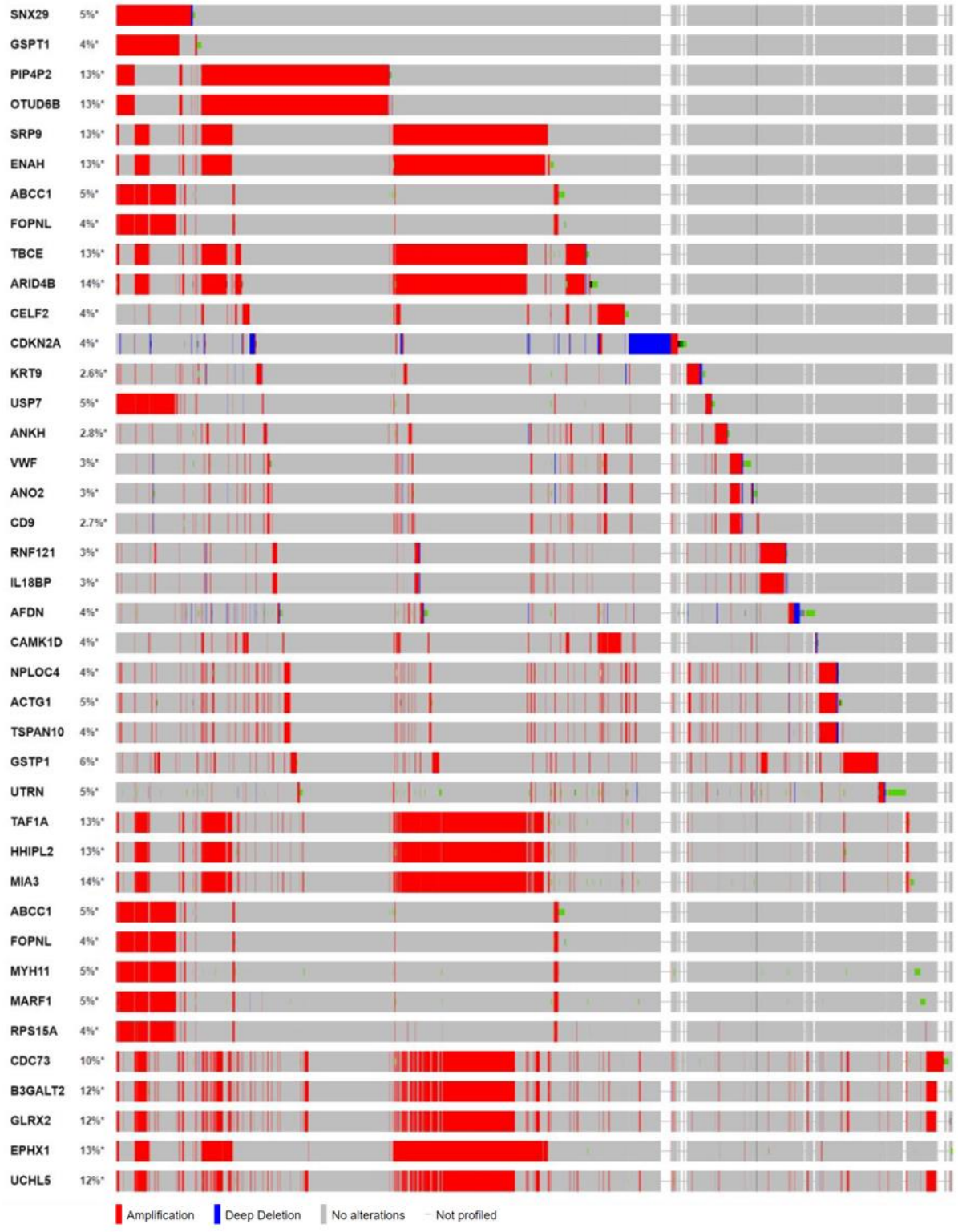
	Chr	Start	End	Length	Fold enrichment	log10 (p-value)	-log10 (q-value)	Located genes	Located region	ENCODE cCREs	Proximate genes
1	chr16	11456403	11456566	351	16.70268	44.57899	35.14319	Snx29	Intron	Distal enhancer	Gspt1
2	chr16	94235831	94236061	231	12.83027	30.79792	23.13297	Hlcs	Intron		Ripply3
3	chr3	56239818	56239976	206	12.00047	28.18617	20.68988		IGR		Nbea , GM25727
4	chr3	107823297	107823609	204	9.63238	27.85611	20.42463	Gm36211	Intron	Distal enhancer	Csf1, Eps813
5	chr4	14898860	14899070	194	11.72387	26.90069	19.48192	Pip4p2	Intron	Distal enhancer	Otud6b, Gm11837
6	chr9	103670753	103671075	193	10.91858	26.62870	19.32	Tmem108	Intron		Bfsp2, Gm38032
7	chr3	146655081	146655266	191	10.83192	26.47301	19.17393	Samd13	Intron	Distal enhancer	Gm16325, 4930503B30 Rik
8	chr11	100198774	100198898	183	10.43743	25.42371	18.34202		IGR		Krt9
9	chr19	40675408	40675578	173	10.89407	24.37279	17.35529	Entpd1	Intron		Cc2d2b, Tctn3
10	chr5	14073258	14073402	162	10.61747	23.13135	16.21274	Gm4128, Sema3e	Intron		
11	chr8	75089607	75089818	151	9.47514	21.89293	15.10229		IGR	Distal enhancer	Hmox1, Mcm5
12	chr19	9978532	9978675	144	9.83882	21.06359	14.44668		IGR	Distal enhancer	Fth1, Rab3il1
13	chr3	121577173	121577280	142	8.40401	20.88354	14.27381		IGR	Distal enhancer	Slc44a3
14	chr9	122471682	122471850	141	9.23456	20.74566	14.13852		IGR		Gm35454, A730085K0 8Rik
15	chr1	45860464	45860635	141	9.78767	20.69623	14.10862		IGR	Distal enhancer	Slc40a1, Wdr75
16	chr1	182071986	182072166	135	9.02538	20.09868	13.5751		IGR		Srp9, Enah
17	chr16	14352755	14352880	130	9.23447	19.50378	13.02702		IGR	Distal enhancer	Abcc1, Fopnl
18	chr17	4626208	4626482	130	9.51107	19.50378	13.02702		IGR		4930517M0 8Rik
19	chr11	75383549	75383810	129	7.64764	19.36886	12.97207	Smyd4	Intron	Distal enhancer	Rpa1, Mir22hg, Prpf8
20	chr19	46227506	46227635	116	7.38092	17.93685	11.63009	Gbf1	Intron	Distal enhancer	Pitx3, Nfkb2
21	chr14	30564535	30564657	113	5.94044	17.65711	11.36401	Tkt	Intron	Distal enhancer	Dcp1a, Prkcd
22	chr13	14034034	14034181	109	8.68126	17.17250	10.91453	Tbce	Intron	Distal enhancer	Arid4b
23	chr13	34398978	34399081	109	8.68126	17.17250	10.91453		IGR	Distal enhancer	Slc22a23
24	chr19	43771687	43771865	109	8.68126	17.17250	10.91453		IGR	Distal enhancer	Abcc2, Cox15
25	chr2	6965357	6965493	109	8.68126	17.17250	10.91453	Celf2	Intron		
26	chr4	89351638	89351756	109	8.68126	17.17250	10.91453	Gm12610	Intron	Distal enhancer	Cdkn2a, Gm12609
27	chr8	75083921	75084062	106	7.97455	16.87909	10.66024		IGR	Distal enhancer	Hmox1, Mcm5
28	chr8	60859669	60859799	106	8.47741	16.83304	10.61735		IGR	Distal enhancer	Hpf1, Nek1
29	chr2	118659235	118659329	101	7.82032	16.37119	10.17369		IGR	Distal enhancer	Pak6
30	chr8	107139891	107140000	101	7.79112	16.31975	10.12425		IGR	Distal enhancer	Cyb5b, Terf2, Cog8, Nfat5
31	chr9	119641996	119642170	101	7.79112	16.31975	10.12425	Scn10a	Intron	Distal enhancer	Scn5a
32	chr15	97786821	97786915	98	6.11812	16.03933	9.86455	Slc48a1	Intron	Distal enhancer	Rapgef3, Endou
33	chr1	40033898	40034006	98	8.12806	16.03524	9.86307		IGR	Distal enhancer	4930456G14 Rik, Il1r2
34	chr15	27219486	27219587	98	8.12806	16.03524	9.86307		IGR		Ankh, Fbx17
35	chr16	9132257	9132367	98	8.12806	16.03524	9.86307	Gm49448	Intron	Distal enhancer	Gm49451, Usp7
36	chr2	112433536	112433663	98	8.12806	16.03524	9.86307			Distal enhancer	Emc7, Aven, Nutm1
37	chr6	125672011	125672107	98	8.12806	16.03524	9.86307	Vwf	Intron		Ano2, Cd9

38	chr8	78110402	78110526	98	8.40466	16.03524	9.86307		IGR	Distal enhancer	Tct29, Slc10a7, Ednra
39	chr4	132624376	132624530	91	7.55055	15.25735	9.18591		IGR	Distal enhancer	Eya3, Atpif1, Sesn2, Rpa2
40	chr2	34869758	34869857	88	7.85146	14.91803	8.86747	Psmd5	Intron	Proximal enhancer	Cutal
41	chr7	40388299	40388410	88	7.85146	14.91803	8.86747	Gm28807	Intron		Gm24910, 4933402C06 Rik
42	chr7	102025872	102025976	88	7.85146	14.91803	8.86747	Rnf121	Intron	Distal enhancer	Il18bp, Xndc1
43	chr3	108013341	108013442	84	6.65687	14.46726	8.48032	Gstm1	Intron	Distal enhancer	Gstm2, Gstm3
44	chr2	18019235	18019358	84	7.07864	14.42945	8.44819		IGR		H2a12a, Gm13330, Mllt10
45	chr3	69118880	69118997	81	7.25196	14.11521	8.15494	Kpna4	Intron	Distal enhancer	Gm1647, Trim59, Ift80
46	chr6	88779880	88780004	79	6.87328	13.86720	7.91727	Mg11	Intron		Tpra1, Kbtbd12
47	chr10	92878077	92878238	78	7.29826	13.82184	7.87402	Cfap54	Intron		Nedd1, Cdk17
48	chr12	98522216	98522319	78	7.57486	13.82184	7.87402	Kcnk10	Intron		Gm40893, Spata7
49	chr17	13905736	13905851	78	7.57486	13.82184	7.87402	Afdn	Exon		Gm7168, Tcte2, Gm49936
50	chr2	5615054	5615146	78	7.57486	13.82184	7.87402	Camk1d	Intron	Distal enhancer	Nudt5
51	chr6	49928212	49928314	78	7.29826	13.82184	7.87402		IGR		Gm3455, Mpp6
52	chr7	75587918	75588015	78	7.29826	13.82184	7.87402	Akap13	Intron	Distal enhancer	Klhl25, Sv2b
53	chr3	102270955	102271073	78	7.12722	13.70250	7.83219	A230001 M10Rik	Intron		Vangl1, Ngf
54	chr11	120387328	120387445	77	6.20420	13.58947	7.72104	Nploc4	Intron	Distal enhancer	Faap100, Actg1, Tspan10
55	chr3	137879440	137879568	70	6.56427	12.91970	7.0817	Dnajb14	Intron	Proximal enhancer	Gm40155, Lamtor3
56	chr19	4046046	4046151	70	6.65837	12.89518	7.05795		IGR	Distal enhancer	Gstp2, Gstp1
57	chr10	119757395	119757485	69	7.02166	12.74773	6.91906	Grip1, Grip1os2	Intron		Tmbim4
58	chr14	99548102	99548269	69	7.02166	12.74773	6.91906	Gm41230	Intron		Dis3, Mzt1
59	chr18	50498901	50499018	69	7.02166	12.74773	6.91906		IGR	Distal enhancer	Prr16
60	chr6	55014367	55014455	69	7.02166	12.74773	6.91906		IGR		Ggct, Gars, Nod1
61	chr4	89234546	89234669	69	6.54604	12.69885	6.90955		IGR		Gm12602, Gm12609
62	chr10	12299053	12299148	68	6.60483	12.63720	6.85258		IGR	Distal enhancer	Utrn,Gm48722, Gm48721
63	chr13	66300097	66300210	65	6.03054	12.33110	6.55919	Gm26754	Intron		AC192334.1
64	chr5	136109233	136109341	65	6.03054	12.33110	6.55919	Rasa4	Intron		Polr2j, Alkbh4
65	chr1	183387785	183387898	64	6.34768	12.18861	6.42096		IGR		Taf1a, Hhipl2, Mia3
66	chr10	119727123	119727212	59	6.74506	11.69687	5.97219	Grip1	Intron		Gm47027, Grip1os3
67	chr11	82174984	82175081	59	6.74506	11.69687	5.97219		IGR	DNase-H3K4me3	Ccl1, Gm17268
68	chr16	14350870	14351017	59	6.74506	11.69687	5.97219		IGR	Distal enhancer	Abcc1, Fopnl, Myh11, Marf1
69	chr16	55616570	55616663	59	6.74506	11.69687	5.97219		IGR		Gm23003, Zpld1
70	chr9	23168286	23168415	59	6.74506	11.69687	5.97219		IGR	Distal enhancer	Bmper
71	chr6	113487411	113487530	59	5.31599	11.60728	5.92191	Creld1	Intron	Distal enhancer	Fancd2, Cidec, Rpusd3
72	chr16	94060214	94060314	57	6.24023	11.43644	5.75826		IGR		Sim2
73	chr7	30117984	30118090	52	5.30205	10.92847	5.28597		IGR		Zfp382, Zfp566, Zfp82, Zfp14, Cox7a1
74	chr16	94956613	94956707	50	6.19186	10.67058	5.04474	Kcnj6	Intron		Kcnj6 (splice variant), Kcnj15

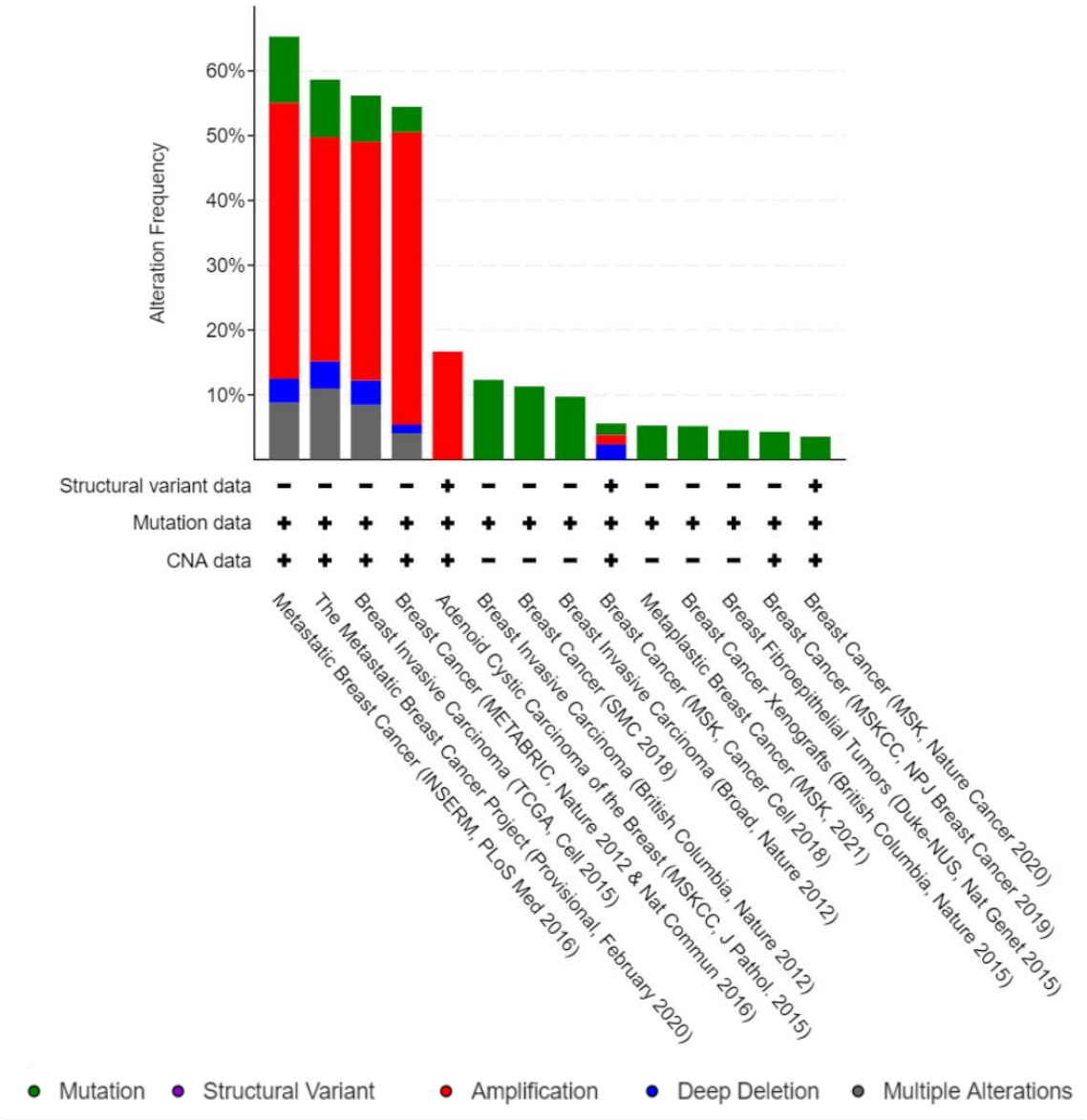
75	chr18	23774354	23774490	50	6.19186	10.67058	5.04474	Mapre2	Intron	Distal enhancer	Gm15972, Mapre2 (splice variant)
76	chr3	104322747	104322845	50	6.19186	10.67058	5.04474		IGR		Magi3, Slc16a1
77	chr3	123256889	123256987	50	6.46846	10.67058	5.04474		IGR	Distal enhancer	Sec24d, Synpo2
78	chr3	126811766	126811927	50	6.19186	10.67058	5.04474	Camk2d	Intron		Ank2, Arsj
79	chr7	118104821	118104931	50	6.46846	10.67058	5.04474	Rps15a	Exon	Distal enhancer	Arl6ip1
80	chr7	27383360	27383499	47	5.86653	10.25134	4.72387	Sptbn4	Intron	Distal enhancer	Gm20479, Shkbp1, Ltbp4, Blvrb
81	chr19	36451218	36451349	47	4.67467	10.22378	4.70391	Pcgf5	Intron	Proximal enhancer	Hectd2
82	chr1	143632998	143633086	41	5.91526	9.67029	4.18256	Cdc73	Intron		B3galt2, Glrx2, Uchl5
83	chr18	81253058	81253164	41	5.91526	9.67029	4.18256		IGR	Distal enhancer	Gm30192, 4930594M17Rik
84	chr7	70398765	70398853	41	5.91526	9.67029	4.18256	B130024G19Rik	Intron		Gm28258, B130024G19Rik, Nr2f2, Gm29683
85	chr3	152677239	152677334	38	5.62596	9.31879	3.89843	Al115009	Intron		Pigk, Ak5
86	chr6	116514979	116515073	38	4.94119	9.25272	3.85098	Olfr212	Intron		Olfr2l3, Gm5580, Olfr214
87	chr12	8076194	8076304	33	5.63866	8.69764	3.33797	Gm33037	Intron	Distal enhancer	Ldah
88	chr1	181009829	181009936	31	3.89556	8.40735	3.11645	Ephx1	Intron	Distal enhancer	9130409I23Rik
89	chr3	51670874	51670974	28	5.29691	8.16043	2.8885	Mgst2	Intron	Proximal enhancer	Setd7
90	chr2	31128124	31128212	25	5.36206	7.75449	2.50403	Fnbp1	Intron	Distal enhancer	Gpr107, Ncs1
91	chr3	60707283	60707373	24	4.90424	7.53305	2.4093		IGR	Distal enhancer	Gm38326, P2ry1
92	chr8	94262758	94262846	24	4.90424	7.53305	2.4093	Nup93	Intron	Distal enhancer	Nup93 (splice variant), Slc12a3, Herpud1
93	chr12	115334193	115334333	18	4.03904	6.90630	1.84058		IGR		Ighv1-58, Ighv8-8
94	chr12	113625529	113625633	16	4.66366	6.68330	1.69745	Ighv5-6	Exon		Ighv2-3, Ighv5-4, Ighv2-2, Ighv5-8



Supplementary Figure 5. One significant peak was detected in the anti-NRF2 antibody (EP1808Y)-enriched 66cl4 NRF2 KD 3B7 sample. A single enriched peak was detected in the negative control sample (third lane) indicating the KD efficiency of NRF2 in the 66cl4 NRF2 KD 3B7 (3B7) cell line. From the top in a consecutive order, each lane presents ChIP-seq data from the following samples: Normal rabbit IgG-precipitated 3B7, input of 3B7, EP1808Y-precipitated 3B7, normal rabbit IgG-precipitated 66cl4 NT (NT), input of NT, EP1808Y-precipiated NT. Each sample exhibited a peak height of 23, 10, 157, 13, 10 and 40, respectively. The ChIP-seq data is displayed with UCSC genome browser [126].



Supplementary Figure 6. Thirty-five genes were highly mutated in clinical breast cancer patients. Among the genes found adjacent or in vicinity of the 94 significant ChIP-seq peaks, 35 were each found to be mutated in more than 3 % of breast cancer patient cohorts according to 16 combined clinical breast cancer studies. Alteration rates of 14 genes were especially surpassing beyond 10 % where ten of the genes have not been identified as NRF2-regulated genes as yet. The red bars indicate gene amplification alterations. The datasets were accessed through cBioPortal [108, 109].



Supplementary Figure 7. Known and novel NRF2-regulated genes of 66cl4 cells may assist the progression of metastatic breast cancer. A considerable number of breast cancer patients with genetic alterations amongst the highly mutated 35 NRF2-regulated genes (known and potential) are diagnosed with the metastatic subtype. Thirty-five genes proximate to the significant ChIP-seq peaks were each altered in more than 3 % of breast cancer patient cohorts according to 16 clinical breast cancer studies. The genetic modification frequency was explicitly higher in metastatic breast cancer studies. The combined datasets and bar graph were obtained from cBioPortal [108, 109].

7.2 Peak caller analyzation of pooled 66cl4 NT input samples

Peak caller: MACS version 2.1.1.20160309 [101]

Command line: callpeak -f BAM -t IP_pooled.bam -c Input_pooled.bam -g 1870000000 -n pooled_IP-INput

ARGUMENTS LIST:

format = BAM

effective genome size = 1.87e+09

band width = 300

model fold = [5, 50]

qvalue cutoff = 5.00e-02

Larger dataset will be scaled towards smaller dataset.

Range for calculating regional lambda is: 1000 bps and 10000 bps

Broad region calling is off

Paired-End mode is off

tag size is determined as 86 bps

total tags in treatment: 60198312

tags after filtering in treatment: 47382114

maximum duplicate tags at the same position in treatment = 1

Redundant rate in treatment: 0.21

total tags in control: 51530051

tags after filtering in control: 34196682

maximum duplicate tags at the same position in control = 1

Redundant rate in control: 0.34

d = 88

alternative fragment length(s) may be 88,510,575 bps

7.3 Material solutions

Cell culture media
- DMEM (500 ml)
- 10% FBS (50 ml)
- 1% L-glutamine (5 ml)
- 1% Penicillin-streptomycin (5 ml)

ChIP-qPCR master mix
- 2X SYBR Green PCR Master Mix (10 µl)
- 2.5 µM Forward primer (2 µl)
- 2.5 µM Reverse primer (2 µl)
- DNA template (2 µl)
- RNase-free water (4 µl)

LiCl wash buffer (cold)
- 250 mM LiCl
- 10 mM Tris pH 8.0
- 1 mM EDTA
- 0.5% NP-40
- 0.5 % Na-Deoxycholate

1.25 M Glycine (cold)
- 75.07 g/mol Glycine (9.38 g)
- Deionized distilled water (100 ml)

10% Na-Deoxycholate
- 2.5 g Na-Deoxycholate
- Deionized distilled water (25 ml)
- Filter sterilization

Protease inhibitor cocktail (complete)
- 50X stock

RIPA buffer (50 ml)
- 10mM Tris pH 8.0
- 1 mM EDTA
- 140 mM NaCl
- 1% Triton X-100
- 0.1% SDS
- 0.1 % Na-Deoxycholate

RIPA/Glycine buffer

10mM Tris pH 8.0
1 mM EDTA
140 mM NaCl
1% Triton X-100
0.1% SDS
0.1 % Na-Deoxycholate
125 mM Glycine

TBS

20 mM Tris-base
137 mM NaCl

TBST

20 mM Tris-base
137 mM NaCl
0.1% Tween 20

TE buffer (cold)

10 mM Tris-HCl pH 8.0
1 mM EDTA

Yuna Shin

Identification of NRF2 binding sites in metastatic breast cancer cells by ChIP-Seq

NTNU
Norwegian University of
Science and Technology