

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Diploma Nutrition Scientist Maria Pudenz

born in Heilbad Heiligenstadt, Germany

Oral-examination: 23rd March 2016

Impact of soy isoflavones on DNA methylation in rat mammary glands

- Dose-response aspects, critical time windows,
and influence on carcinogenesis

Referees: PD. Dr. Odilia Popanda
Prof. Dr. Günter Vollmer

"Life begins at the end of your comfort zone."

Neale Donald Walsch

Contributions

Sections of the introduction are partially based on the review "Impact of Soy Isoflavones on the Epigenome in Cancer Prevention" by Pudenz, M., Roth, K., and Gerhäuser, C. (2014). *Nutrients* 6, 4218-4272 as well as "Genome-Wide DNA Methylation Profiling in Dietary Intervention Studies: a User's Perspective." published by Gerhäuser, C., Heilmann, K., and Pudenz, M. (2015). *Current Pharmacology Reports* 1, 31-45.

Parts of section 4.1 are built on the publication "Dose-dependent effects of isoflavone exposure during early lifetime on the rat mammary gland: Studies on estrogen sensitivity, isoflavone metabolism, and DNA methylation" published by Blei, T.*, Soukup, S.T.*, Schmalbach, K.*, Pudenz, M.*, Möller, F.J., Egert, B., Wortz, N., Kurat, A., Müller, D., Vollmer, G., Gerhäuser, C., Lehmann, L., Kulling, S.E., Diel, P. (2015) *Molecular Nutrition and Food Research* 59, 270-283 (*shared first authorship).

Parts of section 4.3 are built on the publication "Soy-isoflavone exposure through all life stages accelerates E2-induced mammary tumor onset and growth, yet reduces tumor burden, in ACI rats" Möller, F.J., Pemp, D., Soukup, S.T., Wende, K., Zhang, X., Zierau, O., Muders, M.H., Bosland, M.C., Kulling, S.E., Lehmann, L., Vollmer, G. (2015) *Archives of Toxicology* (in revision).

Therefore, parts of the text in these sections might contain suggestions and corrections from co-authors.

Parts of section 1.2.4 and the methods section 3.2.5 was built on the master thesis: "Impact of Soy Isoflavones on Expression of miRNA and their targets during estrogen-induced mammary carcinogenesis in ACI rats" presented by Kevin Roth to the Faculty of Biosciences at the Ruperto-Carola University of Heidelberg in 2015.

The statistical analyses (except for Two-way ANOVA) was performed by Dr. Thomas Hielscher (Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany) the methodological description included in section 3.2.12 was written by him.

The development of the scripts for alignment and quality control of the MClp-Seq (section 3.2.6.3) was done by Dr. Lei Gu (Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center, Heidelberg, Germany. Present addresses: Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, USA). Dr. Daniel Lipka (Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center, Heidelberg, Germany) provided significant support by the application of the Homer suite of tools. David Brocks (Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center, Heidelberg, Germany) was involved in the generation of Figure 13 using the MEDIPS package of R.

RRBS library preparation was implemented in close cooperation with Christoph Bock (CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria). Read alignment and initial bioinformatics analyses of the sequencing data were also performed at CeMM. After initial bioinformatics, RnBeads analysis (section 3.2.7.7) was done by Dr. Yassen Assenov and later by Dr. Clarissa Gerhäuser (Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center, Heidelberg, Germany).

Mammary gland samples of Wistar rats and the corresponding data of biometric parameters and IHC were provided by Prof Dr. Dr. Patrick Diel, on behalf of the Institute of Cardiovascular Research and Sports Medicine, Department of Molecular and Cellular Sports Medicine, German Sport University, Cologne, Germany.

Mammary gland samples of ACI rats and the corresponding data of biometric parameters and IHC were provided by Prof Dr. Günter Vollmer, on behalf of the Department of Molecular Cell Physiology and Endocrinology, Technical University Dresden.

Declaration

Declarations according to § 8 (3) b) and c) of the doctoral degree regulations:

a) I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those expressly indicated,

b) I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

(Maria Pudenz)

List of contents

CONTRIBUTIONS.....	III
DECLARATION.....	V
SUMMARY.....	XIII
ZUSAMMENFASSUNG.....	XV
LIST OF ABBREVIATIONS.....	XIX
1. INTRODUCTION.....	1
1.1 ISOFLAVONES.....	1
1.1.1 Sources, chemistry and dietary intake.....	1
1.1.2 Bioavailability of IF.....	2
1.1.2.1 Absorption and excretion.....	2
1.1.2.2 Metabolism and distribution.....	2
1.1.3 Biological effects.....	3
1.1.3.1 ER-mediated dualistic mode of action.....	3
1.1.3.2 E2-independant mechanisms of action.....	4
1.1.3.3 Specific disease endpoints.....	5
Menopausal symptoms.....	5
Bone health.....	6
Cardiovascular effects.....	6
1.2 EPIGENETIC REGULATION OF GENE EXPRESSION.....	8
1.2.1 DNA methylation.....	8
1.2.1.1 Plasticity of DNA methylation.....	8
1.2.1.2 Enzymes catalyzing DNA methylation and demethylation.....	9
Writers and erasers.....	9
Readers.....	10
1.2.2 Methods to detect DNA methylation.....	11
1.2.2.1 Genome-wide methylation profiling.....	11
1.2.2.2 Methyl-CpG immunoprecipitation-sequencing (MCIp-Seq).....	13
1.2.2.3 Reduced representation bisulfite sequencing (RRBS).....	14
1.2.2.4 EpiTYPER MassARRAY.....	14
1.2.3 Histone tail modifications.....	16
1.2.4 Non coding RNA (ncRNA).....	17
1.3 MAMMARY GLANDS, BREAST CANCER AND PREVENTIVE IMPACT OF IF.....	18
1.3.1 Normal mammary gland development.....	18
1.3.1.1 Childhood, puberty and adolescence.....	18
1.3.1.2 Pregnancy, lactation and menopause.....	19
1.3.2 Breast cancer development.....	19
1.3.2.1 Breast cancer progression.....	20

1.3.2.2	Breast cancer classification	21
1.3.2.3	Epidemiology of breast cancer and the impact of IF	22
1.3.2.4	DNA methylation changes associated with breast cancer	23
1.3.2.5	Impact of IF on DNA methylation in breast cancer	24
1.4	ISOCROSS PROJECT	26
2.	AIMS.....	27
3.	MATERIAL AND METHODS.....	29
3.1	MATERIAL.....	29
3.1.1	<i>Instrumentation</i>	29
3.1.2	<i>Software and databases</i>	29
3.1.3	<i>Primers</i>	30
3.1.3.1	MassARRAY primers	30
3.1.3.2	Primers for genome-wide methylation profiling	32
3.1.3.3	Roche UPL qPCR primers	33
3.2	METHODS.....	35
3.2.1	<i>Experimental diets, animal subjects and study design</i>	35
3.2.2	<i>Soy isoflavone enriched diets</i>	35
3.2.3	<i>Rat uterotrophy assay in Wistar rats</i>	36
3.2.4	<i>August Copenhagen Irish rats</i>	37
3.2.5	<i>DNA, RNA and protein isolation</i>	41
3.2.5.1	Tissue processing.....	41
3.2.5.2	DNA isolation.....	42
3.2.5.3	RNA isolation	42
3.2.5.4	Protein isolation	43
3.2.6	<i>Methyl-CpG Immunoprecipitation and Next generation Sequencing</i>	43
3.2.6.1	High salt and single fraction elution protocol.....	44
3.2.6.2	Control for enrichment of methylated DNA by qPCR	45
3.2.6.3	Bioinformatics data mining	45
3.2.7	<i>Reduced representation bisulfite sequencing - Library preparation</i>	46
3.2.7.1	MspI digest, end repair and A-Tailing.....	46
3.2.7.2	Adapter Ligation	47
3.2.7.3	Cleaning up using SPRI Beads	47
3.2.7.4	Quality control qPCR	48
3.2.7.5	Bisulfite conversion	49
3.2.7.6	Enrichment PCR.....	49
3.2.7.7	Bioinformatics analysis.....	50
3.2.8	<i>Quantitative DNA Methylation analysis by EpiTYPER technology</i>	51
3.2.8.1	Bisulfite conversion	52
3.2.8.2	Primer design and optimization	53
3.2.8.3	Preparation of methylated and unmethylated standard.....	53
3.2.8.4	PCR amplification	55

3.2.8.5	SAP treatment, <i>in vitro</i> transcription and desalting	55
3.2.8.6	Analysis of quantitative methylation data.....	56
3.2.9	<i>Quantification of nucleic acids using real time quantitative PCR (RT-qPCR)</i>	56
3.2.9.1	cDNA synthesis	56
3.2.9.2	Real time quantitative PCR.....	57
3.2.10	<i>Laser capture microdissection (LCM)</i>	57
3.2.11	<i>Immunohistochemistry (IHC)</i>	59
3.2.12	<i>Statistical analysis</i>	59
3.2.12.1	Wistar rats and developmental effects in ACI rats	59
3.2.12.2	E2-induced breast carcinogenesis in ACI rats	59
4.	RESULTS	61
4.1	EFFECTS OF IF EXPOSURE ON HORMONE RESPONSIVITY IN HEALTHY WISTAR RATS.....	61
4.1.1	<i>Onset of puberty and rat uterotrophic assay</i>	62
4.1.2	<i>Estrogenic and proliferative response measured by reference genes</i>	63
4.1.3	<i>IHC in mammary gland tissue</i>	65
4.2	DNA METHYLOME MINING IN NORMAL MAMMARY GLANDS OF WISTAR RATS	67
4.2.1	<i>Enrichment of high and medium methylated DNA for MClp-Seq</i>	67
4.2.2	<i>Identification of IF-induced differential methylated regions</i>	68
4.2.3	<i>Genomic coverage</i>	68
4.2.3.1	Relative genomic distribution.....	69
4.2.3.2	Overlap and TF motif enrichment	69
4.2.4	<i>Confirmation of IF-induced DMRs obtained from MClp-Seq using MassARRAY</i>	71
4.2.5	<i>Response to dietary IF is dependent on dose and timing of exposure</i>	72
4.2.5.1	Critical time windows	72
4.2.5.2	Dose aspects of IF intervention	74
4.2.6	<i>Effects of an IF intervention are reflected in gene expression patterns</i>	78
4.3	LIFELONG EXPOSURE TO DIETARY IF AFFECTS SPECIFIC ENDPOINTS IN ACI RATS.....	82
4.3.1	<i>Kinetics of DNA methylation in the course of normal MG development</i>	82
4.3.1.1	Onset of puberty and effects on physiological parameters.....	82
4.3.1.2	Estrogenic and proliferative response of the MG across developmental stages.....	83
4.3.1.3	IHC in mammary gland tissue	85
4.3.2	<i>Implementation of a RRBS library preparation protocol and trouble shooting</i>	85
4.3.2.1	RRBS library preparation	86
4.3.2.2	Library preparation troubleshooting	87
	Loss of fragments.....	87
	Bisulfite conversion rate	88
4.3.3	<i>Epigenetic profiling of developmental effects during puberty</i>	89
4.3.3.1	Technical confirmation of quantitative methylation analyses	91
	Selection of candidate regions for RRBS confirmation	91
	RRBS confirmation using MassARRAY technology	92
4.3.3.2	Quantitative analysis of candidate regions across all developmental windows.....	94

4.3.4	<i>Gene expression analyses</i>	96
4.3.4.1	Candidate gene expression and correlation with DNA methylation changes	96
4.3.4.2	Gene expression-based pre characterization using cell type marker genes.....	97
4.3.4.3	IF affect gene expression of key epigenetic enzymes.....	99
4.3.5	<i>Epigenetic profiling during E2-induced breast carcinogenesis in ACI rats</i>	100
4.3.5.1	E2-induced carcinogenesis in the MG	100
	IF treatment affects various physiological parameters.....	102
	Impact of IF intervention on estrogenic and proliferative response in MGs and E2-induced tumors	102
4.3.5.2	DNA methylome mining during tumor formation	104
	Genomic Distribution	104
	The chemopreventive effects of lifelong IF exposure evolve in distinct biological processes	107
4.3.5.3	Technical confirmation of quantitative methylation levels from RRBS	108
	Selection of candidate regions for RRBS confirmation in adult ACI rats	108
	Technical confirmation of RRBS data using MassARRAY for MG and tumor samples of ACI rats	110
	Correlation of MassARRAY and RRBS data of MGs and tumors of ACI rats	113
4.3.5.4	Quantitative analysis of candidate regions during MG tumor development	114
4.3.6	<i>Gene expression analyses</i>	118
4.3.6.1	Candidate gene expression and correlation with DNA methylation changes	118
4.3.6.2	Gene expression-based cell type characterization using reference markers	121
4.3.6.3	IF affect gene expression of key epigenetic enzymes.....	123
4.3.7	<i>DNA methylation changes of enriched MG epithelial cells obtained by LCM</i>	124
5.	DISCUSSION	127
5.1	METHOD ESTABLISHMENT	127
5.1.1	<i>MClp-Seq in Wistar rats</i>	127
5.1.1.1	HS elution: enriching for highly methylated DNA.....	128
5.1.1.2	SF elution: enriching for moderate and highly methylated DNA.....	128
5.1.1.3	Confirmation of MClp-Seq data.....	130
5.1.2	<i>RRBS in ACI rats</i>	131
5.1.2.1	Genome-wide methylation screen during puberty	132
5.1.2.2	Methylome profiling during E2-induced breast carcinogenesis	133
5.1.2.3	Stringent cutoff criteria for RRBS data secure reliable quantitative methylation data	135
	Results obtained in the puberty study	135
	Results obtained in the carcinogenesis study	136
5.2	BIOLOGICAL EFFECTS	138
5.2.1	<i>Investigations on IF dose and exposure time in Wistar rats</i>	138
5.2.1.1	Specific TF are enriched in DMRs identified by genome-wide analyses	138
5.2.1.2	IF dose aspects	139
	Onset of puberty is accelerated by lifelong exposure to low and high IF doses	140
	Lifelong high but not low IF doses reduce proliferative and estrogenic response.....	142
	Lifelong low IF exposure induces a U-shape like DNA methylation and mRNA expression profile.....	142
5.2.1.3	Critical time windows	144

Exposure during puberty.....	145
Exposure post OVX mimicking menopause.....	147
5.2.2 <i>Investigations on the kinetics of DNA methylation and on carcinogenesis in ACI rats</i>	149
5.2.2.1 Kinetic experiment	149
Endocrinological parameters are not affected by lifelong high IF exposure.....	149
Proliferative and estrogenic response is increased by lifelong high IF exposure in young rats	149
Kinetics of DNA methylation is stronger affected by age than by diet	150
Expression of epigenetic enzymes is modulated by age and IF exposure.....	151
DNA methylation changes reflect alterations in glandular structure.....	152
5.2.2.2 Carcinogenesis experiment	153
IF exposure reduced incidence of E2-induced MG tumors but accelerated onset and growth	154
Estrogenic response is not affected by lifelong high IF exposure in adult rats	154
DNA methylation and gene expression of candidate genes	155
Writers and erasers of DNA methylation	157
Functional pathways affected by lifelong IF exposure	157
5.2.2.3 Quantitative DNA methylation analyses of enriched MG epithelial cells.....	159
6. CONCLUSIONS AND OUTLOOK.....	161
7. REFERENCES	XXIII
8. SUPPLEMENTS	XLI
PUBLICATIONS AND POSTER PRESENTATIONS	LXI
ACKNOWLEDGEMENTS	LXIII

Summary

Epidemiological studies indicate a reduced risk and a lower incidence of hormone-dependent cancer types in populations following a traditional Asian diet rich in phytoestrogens such as isoflavones (IF). Lifestyle and timing of exposure to IF seems to be a critical factor. In rodent models, especially multi-generational or prepubertal exposure was shown to modulate mammary gland (MG) morphology, resulting in anti-tumorigenic activity. Many processes during early development are regulated by epigenetic mechanisms, therefore IF might alter the epigenetic reprogramming of the MG affecting normal cell growth and susceptibility to breast cancer. Several investigations provide evidence for the impact of IF on DNA methylation but most of the work focused on selected candidate genes. Investigations of genome-wide changes of DNA methylation in the MG upon IF treatment are scarce.

The present thesis was aimed to analyze DNA methylation profiles in MGs of healthy rats and during estrogen-induced mammary carcinogenesis on a genome-wide scale. Our focus was to clarify the impact of different IF doses and identify critical developmental windows of exposure in Wistar rats, and to investigate the chemopreventive efficacy of IF during estrogen-induced MG tumor development in ACI rats.

Genome-wide methylation analyses were initially performed with Wistar rats bred on a diet with medium IF levels that were switched to either IF-free or high dose IF diet for two weeks after ovariectomy (OVX). We observed only few IF-related DNA methylation changes (measured by MCIP-Seq technology). This weak epigenetic effect was attributed to the fact that all rats had been exposed to the phytoestrogens during all critical developmental phases until adult life. Most methylation changes were observed outside of classical regulatory regions (promoters, CpG islands or 5' UTRs) but interestingly were enriched for binding sites for ETS-domain and basic leucine zipper-domain containing transcription factors. For validation of the genome-wide results, selected candidate genes were analyzed quantitatively by EpiTYPER MassARRAY technology. Two groups of rats with lifelong exposure (from conception to post OVX) to low and high IF doses were included to consider dose effects. We observed a U-shaped dose-response pattern, with the low IF concentration reducing methylation levels, whereas medium and high doses increased methylation compared to an IF-unexposed control group. The relevance of these changes needs to be further investigated. However, lifelong exposure to high IF levels significantly reduced estrogenic and proliferative response of the MG, whereas lower IF concentrations were not sufficient to induce the beneficial health effects and even provoked opposing effects, *i.e.*, significantly induced PCNA and PR protein expression.

Different from the lifelong exposure to high IF levels, IF intervention exclusively during puberty exhibited only minor effects on DNA methylation and mRNA expression levels of candidate genes.

Also, expression of proliferation markers were not affected. An exception was an increase in *Vdr* and *Gata3* mRNA expression which might sensitize the MG towards enhanced differentiation.

In contrast, short term IF exposure only after OVX affected DNA methylation patterns of candidate regions mostly in an opposed direction when compared to the lifelong IF exposure. Post OVX exposure to IF also induced pro-proliferative as well as pro-estrogenic properties and sensitized the animals towards the estrogen treatment. These results contradict the use of high IF concentrations, *e.g.*, for hormone replacement therapy after menopause.

In order to investigate lifelong IF-mediated effects on estrogen-induced rat mammary carcinogenesis, ACI were exposed lifelong to the highest IF dose. IF intervention reduced incidence and multiplicity (56 %, $p=0.018$) of MG tumors but shortened tumor latency by 5 weeks ($p<0.0001$) and enhanced tumor growth (>2 fold), if tumors escaped the preventive effect. IF intake increased the estrogenic and proliferative response of the MG during puberty when carcinogenesis was induced by exogenous estrogen exposure, and epigenetic modifying enzymes such as DNMT3a and 3b were significantly downregulated. Genome-wide methylation profiling analyzed by quantitative reduced representation bisulfite sequencing (RRBS) again indicated hypo- and hypermethylation mainly outside of classical regulatory regions. Genes with differential promoter methylation induced during carcinogenesis which could be modulated by IF exposure were enriched for biological processes and signaling pathways involved in reproductive tissue homeostasis and endocrine system. After applying stringent selection criteria, we were able to validate selected methylation changes from the RRBS analysis by MassARRAY. In all regions investigated, the carcinogenic process significantly modified methylation levels up to 30 %, inducing both, hypo- and hypermethylation. Interestingly, lifelong intervention with the highest IF dose prevented this carcinogenesis-mediated loss and/or gain in methylation. For half of the selected candidate genes methylation changes significantly correlated with mRNA expression. Interestingly, lifelong high dose IF exposure significantly reduced mRNA levels of DNMT1 in healthy MGs and prevented the estrogen-induced upregulation of DNMT1 during tumor formation.

In conclusion, we identified genome wide DNA methylation changes induced by dietary IF. The impact of IF on DNA methylation is highly dependent on exposure time and IF dose. Lifelong exposure to IF reduced estrogen-induced MG tumor incidence but shorted tumor latency. This phenomenon might be partly explained by the downregulation of epigenetic modifiers such as DNMTs during early MG development as well as during breast carcinogenesis. Additional investigations are required to gain a comprehensive insight on IF-induced epigenetic regulation unraveling the functional mechanisms of biological effects exerted by phytoestrogenic soy IF.

Zusammenfassung

Epidemiologische Studien zeigen, dass das Risiko für hormonabhängige Krebsarten sowie deren Inzidenz in asiatischen Regionen vermindert ist, in denen traditionell Lebensmittel mit einem hohen Gehalt an phytoöstrogenen Isoflavonen (IF) verzehrt werden. Die Lebensweise sowie der Zeitpunkt der Exposition stellen dabei vermutlich besonders kritische Faktoren dar. In Tiermodellen konnte gezeigt werden, dass vor allem eine generationsübergreifende und präpubertäre IF Exposition über Veränderungen in der Morphologie der Brustdrüse zu einer verminderten Tumorentstehung beiträgt. Da die frühe Brustentwicklung durch eine Vielzahl von epigenetischen Prozessen reguliert wird, könnte es sein, dass IF die epigenetische Reprogrammierung der Brust modulieren und so nicht nur das normale Zellwachstum beeinflussen, sondern auch die Empfindlichkeit der Brust gegenüber der Mammakarzinogenese verändern. Verschiedene Untersuchungen konnten bereits zeigen, dass IF Einfluss auf die DNA Methylierung nehmen können. Da in den meisten Studien nur ausgewählte Kandidatengene untersucht wurden, gibt es kaum Erhebungen, die IF-induzierte Veränderungen bezogen auf das gesamte Genom zeigen.

Das Ziel der vorliegenden Arbeit bestand darin, Veränderungen im DNA Methylierungsmuster im gesunden Brustgewebe weiblicher Wistar Ratten sowie während der Östrogen-induzierten Mammakarzinogenese genomweit zu erfassen. Besonderes Augenmerk wurde dabei auf den Einfluss der applizierten Dosis und des Zeitpunktes der Exposition gelegt sowie auf das chemopräventive Potential der IF während der Östrogen-induzierten Karzinogenese in ACI Ratten.

Genomweite Analysen mittels MClp-Seq Technologie wurden zunächst mit Wistar Ratten durchgeführt, die unter Verwendung von IF-haltigem Futter (mittlere Dosis) gezüchtet worden waren. Nach Ovariectomie (OVX, vergleichbar mit der Menopause) bekamen die Tiere entweder IF-freies Futter oder eine Diät mit hoher IF Konzentration. Wir konnten nur geringe Unterschiede im DNA Methylierungsprofil zwischen beiden Gruppen detektieren. Dies lässt sich vermutlich dadurch erklären, dass die Ratten während aller kritischen Entwicklungsphasen bis zum Erwachsenenalter den Phytoöstrogenen ausgesetzt waren. Differentiell methylierte Regionen wurden vornehmlich außerhalb von klassischen Expressions-regulierenden Bereichen (Promoter, CpG Inseln oder 5' UTR) gefunden, aber waren interessanterweise in Bindestellen für Transkriptionsfaktoren mit Leucin-Zipper (bZIP)- und ETS-Domäne angereichert. Für die Validierung der genomweiten Ergebnisse wurden interessante Kandidatengene quantitativ mittels EpiTYPER MassARRAY analysiert. Um eine Dosisabhängigkeit der Methylierungsveränderungen zu beurteilen, wurden zusätzlich zwei Gruppen von Ratten untersucht, die lebenslang (von der Zeugung bis post OVX) niedrigen oder hohen IF Konzentrationen ausgesetzt waren. Wir konnten zeigen, dass im Vergleich zu einer Kontrollgruppe mit IF-freiem Futter die niedrige IF Dosis die DNA Methylierung überwiegend reduzierte, während die

mittlere und hohe IF Konzentration zu erhöhten Methylierungswerten und somit zu einem U-förmigen Dosis-Wirkungsprofil führen. Die Relevanz dieser Ergebnisse muss in weiterführenden Studien untersucht werden. Allerdings zeigte sich, dass bei lebenslanger Exposition mit hoher IF Konzentration die östrogene sowie proliferative Antwort der Brustdrüse signifikant vermindert wurde. Im Gegensatz dazu konnte eine niedrige IF Dosis diese gesundheitlich positiven Effekte nicht induzieren und hatte einen eher negativen Einfluss, der z.B. zu einer signifikant erhöhten PCNA und PR Proteinexpression führte.

Im Unterschied zur lebenslangen Exposition zeigten sich nur geringe Effekte auf die DNA Methylierung und mRNA Expression von Kandidatengenen, wenn die Intervention mit hohen IF-Konzentrationen kurzzeitig und ausschließlich während der Pubertät durchgeführt wurde. Auch das Expressionsmuster der Proliferationsmarker war unverändert. Eine Ausnahme bildet dabei eine erhöhte Genexpression von Vdr und Gata3, welche die Brustdrüse möglicherweise für eine verstärkte Differenzierung sensibilisieren kann.

Interessanterweise führte eine Exposition mit hoher IF Konzentration ausschließlich nach der OVX zu Veränderungen in der DNA Methylierung, die denen der lebenslangen Exposition häufig entgegengesetzt waren. Die post OVX IF Intervention entfaltete außerdem pro-proliferative und pro-östrogene Eigenschaften und verstärkte die Sensibilität der Tiere gegenüber der exogenen Östrogen Gabe. Diese Ergebnisse sprechen gegen die Einnahme von IF Supplementen in hohen Konzentrationen als Hormonersatztherapie nach der Menopause.

Der Einfluss einer lebenslangen Exposition mit hoher IF Konzentration auf die Östrogen-induzierte Brustkarzinogenese wurden in ACI Ratten untersucht. Die Intervention resultierte in einer verminderten Tumor-Inzidenz sowie Multiplizität (56 %, $p=0.018$). Wenn aber Tumoren den präventiven Effekten der IF entgingen, führte die Intervention zu einer signifikant verkürzten Tumor-Latenz (um 5 Wochen) sowie zu einem verstärkten Tumorwachstum (>2 fach). Detaillierte Untersuchungen während des Pubertätszeitraumes zeigten, dass die östrogene und proliferative Antwort der Brustdrüse durch die hohe IF Dosis erhöht wurde, während die Expression der epigenetisch aktiven Enzyme DNMT3a und 3b signifikant durch die lebenslange IF-reiche Diät vermindert wurde.

Genomweite Methylierungsanalysen mittels RRBS deuten erneut darauf hin, dass differentiell methylierte Regionen weitestgehend außerhalb der klassischen Expressions-regulierenden Bereiche nachweisbar sind. Allerdings konnten wir zeigen, dass Gene, bei denen Krebs-induzierte Methylierungsveränderungen in der Promoter Region durch die IF Intervention beeinflusst werden konnten, sich auf eine Vielzahl von Funktionen und Signalwegen in Reproduktionsgewebe sowie im endokrinen System auswirken. Nach Anwendung stringenter Auswahlkriterien konnten wir

Methylierungsveränderungen aus der RRBS Analyse mittels MassARRAY Technologie validieren. Die Östrogen-induzierte Mammakarzinogenese führte zu Hyper- als auch Hypomethylierung bis zu 30 %. Interessanterweise verminderte die lebenslange IF-Hochdosis diese Karzinogenese-vermittelte Zu- und Abnahme der DNA Methylierung in den meisten Fällen. Für die Hälfte der ausgewählten Gene korrelierte die Methylierungsveränderung signifikant mit Expressionsunterschieden, womit der Einfluss der IF auf epigenetische Regulationsmechanismen unterstrichen wird. In Übereinstimmung damit reduzierte die lebenslange IF-Hochdosis die DNMT1 Genexpression in gesunden Brustdrüsen und verhinderte ebenfalls die Östrogen-induzierte Expression während der Tumorentwicklung. Somit konnte abermals gezeigt werden, dass eine lebenslange Intervention mit hohen IF Konzentrationen die Expression epigenetisch aktiver Enzyme und dadurch das DNA Methylierungsmuster beeinflussen kann.

Schlussfolgernd lässt sich feststellen, dass in den Rattenmodellen eine IF Exposition über das Futter zu genomweiten Veränderungen in der DNA Methylierung führte. Wir konnten zeigen, dass die Auswirkungen der IF Intervention stark von der applizierten Dosis sowie vom Expositionszeitraum abhängen. Desweiteren verminderte eine lebenslang IF Hochdosis die Östrogen-induzierte Inzidenz von Brusttumoren, verkürzte allerdings die Tumor-Latenz. Dieses Phänomen könnte dadurch erklärt werden, dass die IF Intervention bereits in der frühen Brustentwicklung sowie während der Mammakarzinogenese Einfluss auf die Expression epigenetisch aktiver Enzyme, wie den DNA Methyltransferasen nimmt. Weiterführende Untersuchungen sind notwendig, um einen umfassenden Einblick in die IF-induzierte epigenetische Regulation zu erhalten. Diese Erkenntnisse können maßgeblich zur Klärung der funktionellen Mechanismen beitragen, die den biologischen Effekten der phytoöstrogen Soja-IF zugrunde liegen.

List of abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ACI	August Copenhagen Irish rat
β -ME	β -mercapto-ethanol
BMD	Bone mineral density
cDNA	Complementary DNA
CGIs	CpG islands
CoA	Coenzyme A
Ct	Cycle threshold
DAI	Daidzein
DCIS	Ductal carcinoma <i>in situ</i>
DNA	Desoxyribonucleic acid
DMR	Differentially methylated region
DMS	Differentially methylated site
DNMT	DNA methyltransferase
E2	17 β -estradiol
EB	Elution buffer
ECM	Extracellular matrix
EDTA	Ethylendiaminetetraacetic acid
EMCA	Estrogen-induced mammary cancer
ER, Esr	Estrogen receptor (ER refers to protein, Esr refers to gene)
GEN	Genistein
GLY	Glycitein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDL	High-density lipoprotein
HMT	Histone methyltransferase
HRT	Hormone replacement therapy
HS	High salt elution
IHC	Immunohistochemistry
IF	Isoflavones
LCM	Laser capture microdissection
LDL	Low-density lipoprotein
lncRNA	Long non-coding RNA
M	Million
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight

MBD-Fc	Methyl-CpG-binding domain-Fc
MCip	Methyl-CpG Immunoprecipitation
MG	Mammary gland
MG/T	Mammary gland without palpable tumor
mRNA	Messenger RNA
miRNA	microRNA
NGS, Seq	Next Generation Sequencing
OR	Odds ratio
OVX	Ovariectomy/Ovariectomized
PND	Post natal day
PPC	PCR primer cocktail
PR, Pgr	Progesterone receptor (PR refers to protein, Pgr refers to gene)
PRC2	Polycomb repressive complex 2
RNA	Ribonucleic acid
RRBS	Reduced representation bisulfite sequencing
RT-qPCR	Real time quantitative PCR
SAP	Shrimp alkaline phosphatase
SF	Single fraction elution
SNP	Single nucleotide polymorphism
T	Tumor
TDLU	Terminal ductal lobular unit
TEB	Terminal end buds
TET	Ten-eleven translocation, methylcytosine dioxygenase
TF	Transcription factor
tRNA	transfer RNA
TSS	Transcription start site
TBE	TRIS/Borate/EDTA buffer
UTR	Untranslated regions
WHI	Women Health Initiative

1. Introduction

1.1 Isoflavones

1.1.1 Sources, chemistry and dietary intake

Isoflavones (IF) are a class of plant estrogens (phytoestrogens) mainly present as glycoside conjugates in the Leguminosae plant family. The most prevalent dietary IF including genistein (GEN), daidzein (DAI), and glycitein (GLY) occur in nutritionally relevant amounts in soybeans and soy-based foodstuffs (Figure 1). Concentrations of total IF (sum of all aglycone equivalents) differ in soy flour (150-210 mg/100 g), Tofu (9-50 mg/100 g), and soy milk (1-10 mg/100 g); IF amounts in soy sauce are generally low (<1 mg/100 g). Nowadays, more processed foodstuffs such as soy cheese (6-25 mg/100 g), soy burgers (6-30 mg/100 g), and meatless soy sausages (3-15 mg/100 g) gain increasing interest and substantially contribute to the daily dietary IF intake [1].

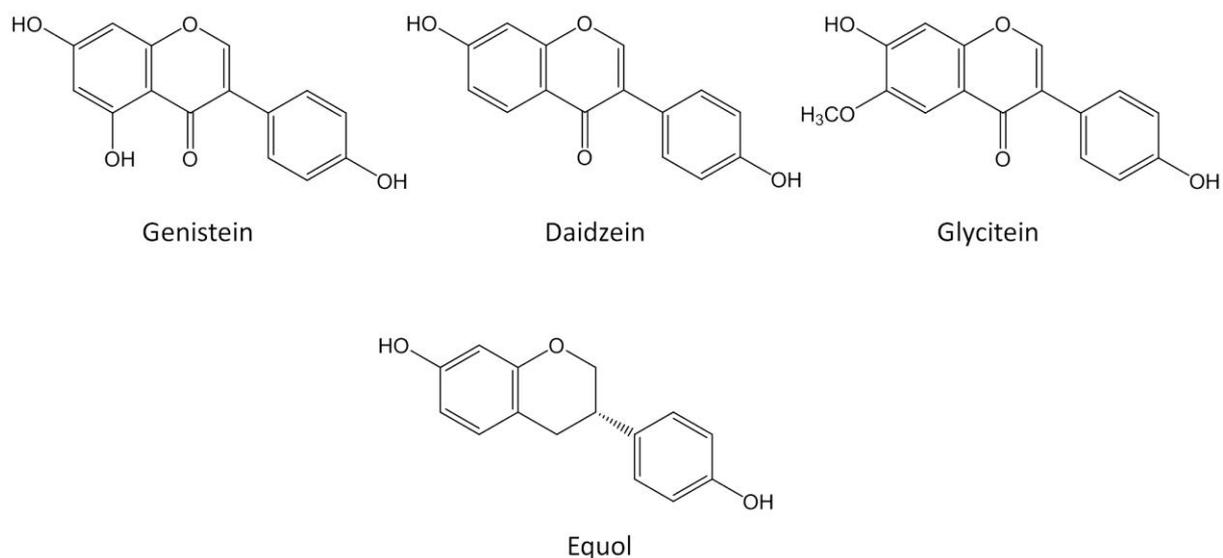


Figure 1: Chemical structures of soy isoflavones.

Chemical structures are depicted as aglycones for predominant isoflavones from soy: Genistein, Daidzein, and Glycitein, as well as the microbial Daidzein metabolite Equol (modified from [2]).

IF levels in soy products vary depending on the soybean cultivar, crop year and weather conditions at time of cultivation. Additionally, processing and storage conditions affect the total IF quantity as well as the IF pattern. Daily dietary intake of IF varies drastically between Western and Asian Countries. Since fermented soy products like Tofu, Tempeh and Miso are part of the regular diet in many Asian

countries, dietary intake of IF has been estimated from 25 mg/day in China [3] and 60 mg/day in Singapore [4] up to 200 mg/day in Japan.

Intakes in the Western society are much lower and do not exceed one to several mg/day [5] with individuals following a vegetarian- (12 mg/day) or vegan-based nutrition (70 mg/day) showing a higher IF intake than the general population. Subgroups with highest IF ingestions are infants fed a soy-based formula (6-47 mg/day), and consumers of soy supplements (50-200 mg/tablet) [6, 7].

1.1.2 Bioavailability of IF

The main requirement for IF to exert biological effects in humans is an effective bioavailability which is determined by their absorption, metabolism, distribution, and excretion.

1.1.2.1 Absorption and excretion

IF are mainly ingested as glycosides which can be hydrolyzed by bacterial or membrane-bound β -glucosidases or β -glucuronidases releasing IF as aglycones that can be either transported across the intestinal wall or get further metabolized by the gut microbiota in the colon. IF show a biphasic pattern of absorption, reflecting constant uptake from the gastrointestinal tract (1-2 h) as well as enterohepatic recirculation (4-6 h) yielding to plasma concentrations of several $\mu\text{mol/L}$ (*e.g.*, after intake of 50 mg IF) [8]. IF metabolites are mainly excreted *via* the urine; half-life values for elimination of IF are >6 h (reviewed in: [9]).

1.1.2.2 Metabolism and distribution

IF undergo considerable biotransformation by the intestinal bacteria. DAI, GEN, and GLY are converted to many specific products *via* demethylation, dehydroxylation, reduction, and ring cleavage. There is extensive interindividual diversity in bacterial metabolism, *i.e.*, 30 % of the adult Western vs. 60 % of the vegetarians and Asian population are able to produce Equol, a DAI metabolite with high binding capacities to the estrogen receptor (ER) (reviewed in: [10]).

Absorbed aglycones are extensively modified by xenobiotic-metabolizing enzymes during Phase I and Phase II metabolism: reduction *via* cytochrome P₄₅₀ enzymes as well as conjugation with glucuronic acid and sulfate. Hence, IF mainly exist as mono- and di-glucuronides, mono- and di-sulfates, and sulfoglucuronides in human plasma which are not able to pass the cell membrane *via* passive diffusion. Specific carriers have been identified that actively transport DAI metabolites (*i.e.*, organic anion transporter (OAT4), organic anion transporting polypeptide (OATP2B), sodium-dependent organic anion transporter (SOAT), Na⁺-taurocholate co-transporting polypeptide (NTCP)) which may also play a role in the cellular uptake of other IF conjugates [11].

It is not well known whether IF metabolites are distributed equally in the human body, and whether IF concentrations in human plasma are a representative marker for concentrations in specific target organs, for example, in breast or adipose tissue. However, it has been reported that consumption of soy products lead to clinically relevant IF levels in mammary glands [8].

1.1.3 Biological effects

In this section, the main observations regarding the impact of IF on physiological and pathological processes will be briefly recapitulated. There is a variety of preclinical *in vitro* and *in vivo* studies investigating the biological effects of IF [12-14]). Numerous reports about transcriptomic, proteomic and metabolomic findings have contributed to describe the complex nature of IF activities (reviewed in: [10, 15-19]).

1.1.3.1 ER-mediated dualistic mode of action

Due to the structural similarity to endogenous 17 β -estradiol (E2), the polyphenolic IF are able to selectively bind to ER, modulating the recruitment of co-repressors and co-activators thus affecting ER-signaling [20].

There are two types of ER: ER α and ER β encoded by *Esr1* and *Esr2*, located on different chromosomes with numerous mRNA splice variants. Upon ligand binding, cytosolic ERs homo and/or heterodimerize and subsequently translocate into the nucleus. Binding to the DNA *via* ER-responsive elements (ERE) recruits co-activators and co-repressors which enables regulation of gene expression of downstream targets. Although ERs share >50 % homology at the ligand binding domain the affinity and response, also to the same ligand, differs and elicits divergent signaling cascades. In addition, the activating function domain (AF1), responsible for recruitment of co-activators and co-repressors, is only conserved by about 20 % between ER α and ER β . Therefore, binding of the same ligand induces different conformational changes. Binding of E2 to ER α favors agonistic effects *via* recruitment of co-activators and favorable antagonistic effects *via* recruitment of co-repressors to ER β . ER α and ER β have opposing effects regarding proliferation, differentiation, and apoptosis and have been shown to regulate one another by forming heterodimers. This interaction influences DNA binding and availability of the respective receptor and results in repressed transcriptional activity [21].

Estrogen-responsive genes are involved in sexual reproduction including development of the mammary gland, reduced cholesterol levels, and plaque production in coronary arteries, as well as increased bone health by balancing bone resorption and bone formation [22]. ERs are expressed in various E2-sensitive tissues, *i.e.*, ER α in endometrium, breast, ovary, and hypothalamic tissue and ER β in kidney, brain, bone, heart, lungs, intestine, prostate, and epithelial cells [23].

GEN shows affinity to both ER α and ER β as the opposing hydroxyl groups at 4'- and 7-position are appropriately positioned to interact with the amino acids in the ER binding pockets in a similar fashion as E2. For a structural comparison of GEN and E2 see Figure 2. However, binding capacity is 15-fold higher for ER β than for ER α as amino acid residues vary between the ER subtypes [24, 25].

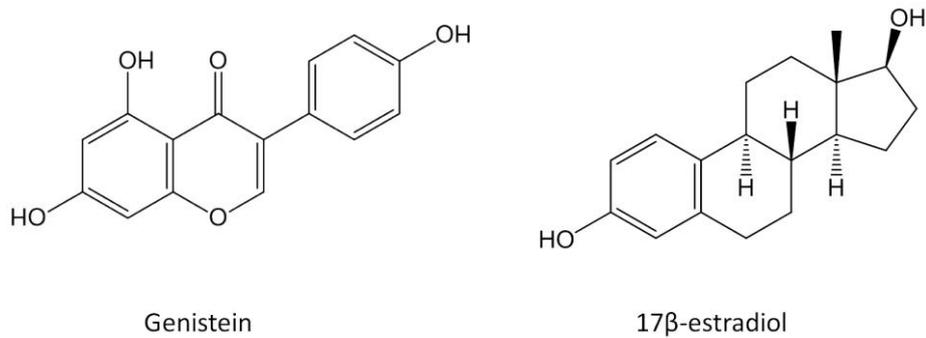


Figure 2: Chemical structures of Genistein in comparison with 17β-estradiol.
Modified from [2])

In principal, IF affinity is approximately 1/1000 of the affinity of E2 to the ER, but their biological effects cannot be ignored since their plasma levels typically exceed circulating endogenous E2 levels by 20-40-fold after consumption of soy-based foodstuffs [8].

IF-binding occurs with distinct affinities in a cell- and tissue-specific manner. Depending on the conformational change of the ER and expression of associated co-regulator proteins IF exert estrogenic effects or anti-estrogenic actions [26]. Since ER are not distributed equally among tissues, IF have been categorized as natural selective estrogen receptor modulators (SERM) [27, 28].

1.1.3.2 E2-independent mechanisms of action

In addition to phytoestrogenic activity, IF display a broad spectrum of mechanisms of actions contributing to cancer preventive potential (reviewed in: [2, 19]). They exhibit antioxidant activity *in vitro* by scavenging free radicals and regulation of enzymes involved in anti-oxidative defense such as catalase and superoxide dismutase thus preventing cells from oxidative stress.

IF also affect synthesis and metabolism of endogenous steroids, *e.g.*, by inhibiting aromatase (CYP19A1), a key enzyme involved in the conversion of testosterone to estrogens. Modulation of xenobiotic metabolism by inhibition of certain Phase I enzymes (*e.g.*, CYP1A1, 1A2, 1B1) and concomitant induction of Phase II enzymes (*e.g.*, GSTs, NQO1, UDPGTs) by IF results in enhanced carcinogen inactivation and detoxification in both *in vitro* and *in vivo* studies.

Cell proliferation is inhibited by IF by regulating cyclin-dependent kinases (CDKs) or their inhibitors (e.g., p21, p16) and promote cell cycle arrest in G2/M. By inducing pro-apoptotic proteins from the BCL2 family, IF treatment was shown to induce apoptosis in numerous studies in cell culture, but also *in vivo*.

IF further inhibit angiogenesis by down-regulating vascular endothelial growth factor (VEGF). Reduced expression of matrix metalloproteinases (MMPs) by IF blocks cell invasiveness and metastasis *in vitro* and *in vivo*.

A number of studies highlight that IF are involved in the modulation of signaling pathways such as epidermal growth factor (EGF) and insulin-like growth factor (IGF-1) signaling, promoting cell differentiation over growth factor-stimulated proliferation and progression.

Further, upregulation of PTEN and inhibition of phosphorylation of I κ B and ERK1/2 abrogated Akt and NF- κ B signaling *in vitro* and *in vivo*. These pathways are known to be involved in sustained cell proliferation and cell survival.

Interestingly, recent evidence, mainly derived from *in vitro* studies, also suggest that IF-mediated impact on epigenetic mechanisms might contribute to their cancer preventive potential (see section 1.3.2.5) (reviewed in [2]).

1.1.3.3 Specific disease endpoints

As described above IF are able to exert complex biological effects through various mechanisms. Perceiving potential beneficial health effects, they are advertised intensively to prevent symptoms associated with the decline of endogenous estrogens (reviewed in [10, 29]).

Menopausal symptoms

With increasing life span, women spend one-third of their lifetime after menopause. Vasomotor symptoms, *i.e.*, hot flushes and night sweats, sleep disturbances, and urogenital atrophy resulting in vaginal discomfort and sexual dysfunction, affect peri- and postmenopausal women and result in reduced quality of life.

For several decades, relief of these symptoms was achieved by prolonged use of hormone replacement therapy (HRT), with concomitant prevention of osteoporosis and occurrence of fractures, without any alarming data published before 2002. With the Women Health Initiative (WHI) trial in 2002, the acceptance of the common HRT treatment regime changed and lead to a fast decline in usage. The WHI investigated 16,608 postmenopausal women aged 50-79 years for 8.5 years on HRT using the most commonly used combined hormone preparation in the US. After 5.2 years of follow-up the WHI trial was stopped due to adverse effects with statistically significant

increased risk for cardiovascular events as well for invasive breast cancer. Since these statistically calculated risks exceeded the benefits observed for prevention of endometrial cancer and osteoporotic fractures, acceptance of HRT was reduced and the application of HRT was stopped [30, 31].

Observation of lower incidence of hot flushes in perimenopausal women in Asian countries (Japan and China) suggested a potential role of IF in reducing menopausal symptoms and shed light on phytoestrogens being an alternative for HRT. Recent systematic reviews propose a slight to modest but significant reduction of the number of daily flushes in menopausal women with a greater benefit for women experiencing a high number of flushes per day [32, 33].

Bone health

So far, epidemiological studies on the impact of IF on bone turnover and mineral density are not fully conclusive. Studies with cultured bone cells and rodent models of postmenopausal osteoporosis clearly support a significant effect of IF and soy protein on bone formation and prevention of bone loss [34, 35]. Kenny et al. did not detect effects on bone mineral density (BMD) after interventions of postmenopausal women with soy protein powder (18 g/d), IF-containing capsules (105 mg/d), or a combination of both for a period of one year [36]. Results were independent of whether individuals were Equol producers or not. In addition, evaluating BMD in women participating in the Shanghai Breast Cancer Survival Study, Baglia et al. illustrated an inverse association of soy intake (mean 48 mg/d) with BMD and an increased odds ratio for osteoporosis in women surviving breast cancer when comparing lowest vs. highest quartiles of IF intake [37].

In contrast, previous reviews clearly stated positive correlations between IF intake and bone health [6, 38]. Ma et al. published a meta-analysis considering all studies available assessing markers for bone resorption (Deoxypyridinoline, Dpyr) in urine and bone formation (Bone-specific alkaline phosphatase, BAP) in serum of pre- and postmenopausal women that ingested soy products or IF for at least 4 weeks. The study discovered that IF intake even below 90 mg/d significantly decreased urinary Dpyr and increased BAP levels, illustrating a positive effect on bone health [39].

Most clinical studies deduced bone turnover from measurement of surrogate markers of osteoblast and osteoclast activity or examined changes in BMD. In order to examine the effects on fracture rates, more long-term studies are needed to provide clear evidence on whether IF affect osteoporosis positively.

Cardiovascular effects

There are numerous epidemiological investigations pointing out the hypocholesterolemic effects of IF leading to reduced low-density lipoprotein (LDL)- and increased high-density lipoprotein (HDL)-

cholesterol levels in human plasma. In a meta-analysis Jenkins et al. assessed 53 studies investigating the potential of soy to reduce LDL-cholesterol in participants independent of gender or menopausal status. Addition of soy protein (50-100 g/d) to the daily diet significantly reduced serum cholesterol by roughly 4 % [40]. Hooper et al. systematically reviewed the effects of different flavonoid products and identified reduced diastolic blood pressure and LDL-cholesterol as significantly affected endpoints after soy protein ingestion, using estimated intakes of 2 mg IF and 25–50 g/d soy protein for studies referring to US and Asian population, respectively [41].

In a recent study by Meléndez et al., cynomolgus monkeys (crab-eating macaque) were exposed to an IF-enriched diet (30 mg/d approximating a human intake of 140 mg /d) or control diet during pre- and postmenopause, or diets were swapped after menopause. Total plasma cholesterol and LDL levels were significantly reduced and HDL concentrations were increased in groups receiving IF diet lifelong as well as after menopause. Interestingly, plaque size in coronary arteries was significantly reduced by lifelong IF exposure, although severity of plaques (AHA grade >III) was also reduced when IF intake was stopped at menopause [42]. In addition, the anti-oxidative potential of IF resulting in reduced LDL oxidation and lipid peroxidation further positively influenced cardiovascular health [43, 44].

Endothelial function (EF) is an independent marker for coronary heart disease and predicts cardiovascular morbidity. In a meta-analysis published by Beavers et al., 17 studies were summarized that determine EF by flow mediated vasodilation (FMD) in postmenopausal women and men after IF supplementation. Seven studies reported a significant increase in FMD, indicating an improved EF beneficially influencing vascular health after >4 weeks of IF exposure (average intake 33-120 mg/d) [45]. These vasodilatory effects might be due to an IF-mediated activation of endothelial nitric oxidation synthase (eNOS) reported by others [46].

In summary, the current literature supports the protective association between IF intake and beneficial health effects regarding cardio vascular disease and reduced menopausal symptoms particularly when women suffer from severely high numbers of hot flushes. Regarding bone maintenance, more research is needed to gain clear evidence for the beneficial influence of IF on bone health and the prevention of osteoporosis, as available data are partly inconsistent. Differences in IF uptake and variability in the gut microbiota composition may influence IF bioavailability as well as gene polymorphisms, *e.g.*, in *Esr1* and *Esr2* may explain discrepancies and individual variability observed in clinical studies investigating biological response to IF.

1.2 Epigenetic regulation of gene expression

The term "epigenetics" was firstly stated in early 1942 by Conrad Hal Waddington and was mostly used to refer to developmental mechanisms in general to describe processes that could not be explained by simply using the terms genotype or phenotype (milestones in epigenetic research reviewed in: [47]).

Nowadays, epigenetics refer to heritable but potentially reversible alterations in gene expression that arise from other than DNA sequence changes. In detail, DNA methylation, histone tail modifications, and non-coding RNAs, as well as nucleosome positioning act coordinately and affect cellular gene expression upon endocrine signals and environmental stimuli such as diet or exposure to xenobiotics [48]. These mechanisms form an epigenetic landscape regulated by numerous enzymes establishing (writers), interpreting (readers), modifying (editors), or removing (erasers) epigenetic marks (reviewed in [49]).

1.2.1 DNA methylation

Holliday was the first to hypothesize that DNA methylation is an epigenetic control mechanism that when lost could lead to changes in genes expression and might be associated with carcinogenesis [50]. This assumption was experimently confirmed by Feinberg and Vogelstein in 1983 who demonstrated that certain oncogenes were significantly hypomethylated in primary human carcinomas when compared to adjacent normal tissue ([51, 52] reviewed in: [47]).

Almost exclusively, DNA methylation is characterized as covalent but reversible addition of a methyl-group to a cytosine base at the carbon-5 position (5-methylcytosine, 5mC). As a consequence of the frequent spontaneous or enzymatic deamination of 5mC to thymine, CG dinucleotides are generally underrepresented in the genome (non-uniform distribution) and are concentrated at distinct genomic regions, so called CpG-island (CGIs). These CGIs often overlap with transcriptional start sites (TSS) and per definition have a minumim length of 200 or 500 bp and show a GC content of 50-55 % [53].

1.2.1.1 Plasticity of DNA methylation

In healthy tissue, promoter CGIs are generally unmethylated and allow for active transcription. About 6-8 % of all promoters are methylated and associated with gene repression maintaining cell-type specific gene transcription. In contrast to methylated CpG-dense regions involved in cellular state and memory preservation, methylation in CpG poor regions, *i.e.*, intragenic and exonic regions is typically connected to active gene expression as well as to alternative splicing. CGI-shores and distal

enhancers are tissue-specificly methylated, whereas intergenic regions, *i.e.*, repetitive sequences and transposable elements are mostly methylated, preserving chromosomal stability. Methylation can modulate gene transcription by blocking transcription factor (TF, *e.g.*, c-Myc) or insulator binding (*e.g.*, CTCF) and, as a consequence, inhibiting transcriptional activation or enabling interaction between distal enhancers and promoters [54, 55].

It is well known that epigenetic patterns undergo intensive alterations, best exemplified by global demethylation in early preimplantation, genome-wide *de novo* methylation after early embryogenesis, and longterm silencing as, for example, with X-chromosome inactivation and imprinting. During aging specific epigenetic pattern are altered by sporadic methylation in CGIs and a global loss of methylation contributing to aberrant gene expression profiles. During carcinogenesis, these alterations are even more pronounced. Global dysregulation of DNA methylation resulting in genome-wide DNA hypomethylation at repetitive sequences and retrotransposons as well as in introns and gene deserts, facilitate genomic instability and promote chromosomal rearrangements. In addition, a decrease in DNA methylation at promoters of proto-oncogenes has been reported as a mechanism of their activation. Conversely, region specific hypermethylation of promoter CGIs of tumor suppressor genes (TSGs) lead to their transcriptional repression equivalent to an inactivating genetic mutation and thus might serve as a second hit according to Alfred Knudson's "two hit" model required for tumor initiation (reviewed in: [56-58]).

1.2.1.2 Enzymes catalyzing DNA methylation and demethylation

Writers and erasers

DNA methyltransferases (DNMTs) are the key enzymes involved in establishing the methylation mark either during early development, replication or differentiation. Three families of DNA (5-cytosine) methyltransferases have been identified. DNMT1 is involved in maintaining methylation patterns with preferences for hemimethylated DNA. Located at the replication fork in a complex together with PCNA, DNMT1 methylates newly biosynthesized DNA. DNMT3a and 3b as *de novo* methyltransferases set up new DNA methylation marks. DNMT3-like (DNMT3L), a catalytically inactive homologue of DNMT3a and 3b does not possess methyltransferase activity, but interacts with DNMT3a and regulates maternal imprinting. DNMT2 rather shows tRNA methyltransferase specificity and is not required for DNA methylation [59]. DNMT3a and DNMT3b are highly expressed in early embryonic cells where most programmed *de novo* methylation occurs and are down regulated in differentiated somatic tissues. All three DNMTs have been reported to be overexpressed in several tumor types such as leukemia and colon, liver, breast and prostate cancer [60]. This upregulation

might serve as an underlying mechanism of the regional hypermethylation of promoter CGIs observed during tumorigenesis, resulting in subsequent silencing of TSGs [61].

The DNA methylation mark can be erased by active and passive processes that work nonexclusively to shape the methylation profiles of a cell. Inhibiting or down-regulating DNMT1 can lead to a passive loss of DNA methylation through DNA replication and to a subsequent reactivation of previously silenced genes. Additionally, re-expression can be achieved by the presence of ten-eleven translocation (TET) methylcytosine dioxygenases 1-3 which lead to active demethylation by an iron- and α -ketoglutarate-dependent sequential oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) and 5-formylcytosine (5fC) to 5-carboxylcytosine (5caC) and subsequent thymine DNA glycosylase (TDG)-mediated base excision repair [62, 63].

Active demethylation predominantly occurs at the paternal genome when transmitted to the zygote leading to rapid loss of paternal epigenomic information. The maternal chromosomes are demethylated during later stages of embryonic development after several rounds of cell divisions in a passive fashion [64]. TET proteins are also involved in maintenance of pluripotency of embryonic stem cells as well as in post-mitotic somatic cells. Particularly, the dynamic response to environmental stimuli is suggesting a role of TET protein in demethylation of individual promoters [65].

Downregulation of TET expression has been observed for various tumor types including breast cancer. Interestingly, leukemia patients with inactivating mutations of TET2 are susceptible to therapies inhibiting the DNMTs [66, 67]. Furthermore, the oncometabolite α -hydroxyglutarate produced by mutant IDH1 and IDH2 inactivates TET enzymatic activity and leads concomitantly to aberrant DNA methylation [68].

Readers

DNA methylation readers, *i.e.*, Methyl-CpG-binding domain (MBD) containing proteins, recognize, interpret, and bind to methylated CpGs. There are four different MBD proteins that bind to methylated DNA and facilitate the binding of the MBD-multiprotein complex 1 (MeCP1). MBD and MeCP1 recruit and interact with co-repressor complexes associated with histone tail-modifying proteins such as histone deacetylases (HDACs) which remove histone acetyl groups by catalyzing the transfer to coenzyme A (CoA). Other co-repressors are histone lysine methyltransferases (HMT) which transfer methyl groups from S-adenosyl-L-methionine (SAM) to lysine residues and induce chromatin condensation and transcriptional repression. In addition, other chromatin remodelers that further promote gene silencing might be recruited [69, 70].

Binding of MBDs to methylated CpGs is favored in highly methylated CpG-dense regions. This characteristic can be used to capture DNA fragments according to their degree of methylation by gradually impede this ionic strength using sodium chloride (see 3.2.6).

1.2.2 Methods to detect DNA methylation

For many years, DNA methylation analysis was restricted to relatively localized regions of the genome or even to single CpG sites [71]. With the development of human microarrays and next generation sequencing (NGS, Seq) now available at affordable costs, several methods are accessible to map DNA methylation on a genome-wide scale. Methods highly vary in the resolution, coverage, and input material that is needed for a reliable and reproducible readout as well as in bioinformatics effort for analysis of the data. Several recent review articles have compared the advantages and limitations of the available methods and summarize considerations for choosing one or the other methodology to answer a certain question of interest or hypothesis [53, 72-75].

1.2.2.1 Genome-wide methylation profiling

Since DNMTs are not present during PCR or in biological cloning systems, DNA methylation information is erased during DNA amplification. Therefore, almost all sequence-specific DNA methylation analysis techniques rely on a methylation-dependent treatment of the DNA before amplification or hybridization [53]. In principal, existing technologies rely on three basic approaches:

i: Digestion with methylation-sensitive restriction endonucleases to fractionate according to the methylation status in methylated and unmethylated DNA fragments. **Methyl-sensitive Restriction Enzyme Sequencing (MRE-Seq/Methyl-Seq)** identifies differential methylation by comparing sequenced fragments from, *e.g.*, MspI and HpaII (the methyl-sensitive isoschizomer of MspI) digests between samples. Information is limited to the recognized cytosine residues, and incomplete digestion might be a source of false results obtained with this approach.

ii: Affinity enrichment of methylated DNA fragments by immunoprecipitation either uses an antibody against 5mC (**methylated DNA Immunoprecipitation (MeDIP)**) or applies the methyl-CpG-binding domain of the human MBD2 protein (**Methyl-CpG Immunoprecipitation (MCIp/MBD capture)**) to reduce the complexity of the whole genome and to enrich for medium and high methylated DNA fragments. MeDIP targets single stranded DNA with regions of lower CpG density. By contrast, MCIp enriches preferentially CpG-dense regions originated from double-stranded DNA. Both enrichment-based technologies do not offer precise quantification of methylation levels as they account for differences in enrichment only. In general, a high nominal coverage of the genome can be achieved but not at single CpG site resolution (resolution >100bp).

iii: Bisulfite treatment uses the advantage of a chemical reaction with sodium bisulfite that deaminates unmethylated cytosine residues to uracil much more rapidly than methylated cytosines and converts the epigenetic methylation mark into a genetic mark [76]. The introduction of a DNA sequence change enables methylation fingerprinting at a single CpG site resolution. However, using bisulfite conversion 5hmC cannot be discriminated from 5mC since 5hmC is not chemically deaminated. Therefore, further manipulation such as oxidation of 5hmC to 5fC which then reacts with sodium bisulfite and is identified as unmethylated C are necessary to achieve a precise readout of 5mC, only [77].

Whole Genome Bisulfite Sequencing (WGBS) is the current gold standard for samples for which a reference genome is available. This technique provides comprehensive genomic coverage as billions of reads per samples are provided at single CpG resolution. WGBS typically requires large amounts of input DNA which might be reduced by the use of a tagmentation-based WGBS (TWGBS) protocol. In this case, genomic DNA is fragmented and tagged with adapters by the enzyme transposase. Intermediate clean-up steps during library preparation are nonessential and input DNA can be reduced to <30 ng [78]. However, WGBS is still extremely costly, sequencing depth is rather low (<20x) and data analysis requires intense downstream computational support.

In order to minimize the fraction of the genome to be sequenced **reduced representation bisulfite sequencing (RRBS)** uses restriction endonucleases that digest the genomic DNA at specific recognition sites. By reducing the complexity to approximately 1 % of genome but still covering 10 % of all available CpGs this technique represents a cost efficient alternative to WGBS and still retains single CpG site resolution.

Illumina arrays are an option apart from sequencing-based methods to profile the human methylome at a number of preselected genomic regions. Two platforms are available, Infinium HumanMethylation27k and 450k BeadChip arrays, which allow the analysis of 27,578 and 482,422 DNA methylation sites, respectively, with up to 12 samples at a time. Bisulfite-treated DNA is hybridized to bead-bound probes representing the methylated or unmethylated condition of a particular locus, and methylation information is detected by fluorescent labels.

iv: Methods without modification of the DNA are available using Single-Molecule sequencing in Real Time (SMRT) instruments and nanopore-based platforms. SMRT sequences single DNA molecules by observing a single strand of DNA being replicated in real time. These third generation technologies are based on nucleotide incorporation times and the different kinetic signals of 5mC, 5hmC, 5fC and 5caC [79]. However, error rates in base identification are still high and feasibility to analyze mammalian size genomes has to be improved for routine analysis.

1.2.2.2 Methyl-CpG immunoprecipitation-sequencing (MCIp-Seq)

MCIp is one of the affinity enrichment-based methods for genome-wide methylation profiling. The MCIp assay enables sensitive detection of CGI-specific methylation using a recombinant bivalent methyl-CpG-binding polypeptide that binds methylated DNA fragments and allows fractionated binding and subsequent detection of hypermethylation on a genome-wide scale [80, 81].

The antibody-like fusion protein MBD-Fc: methyl-CpG-binding domain of the human MBD2 protein and the Fc portion of the human IgG1 show very high affinity to double-stranded methylated DNA. Affinity is dependent on (i) the amount of salt in washing buffer where higher salt concentrations require more 5mC for binding, (ii) number and degree of 5mC in a fragment and (iii) MBD-Fc protein density (Figure 3) [81]. MBD2 shows the highest affinity to methylated DNA when compared to other MBD proteins, *e.g.*, MeCP2, allowing a smaller scale DNA input with less bias towards certain CpG motifs [82].

Adjusting the amount of salt in the washing buffer enables the enrichment of DNA fragments according to their degree of methylation since affinity to the MDB protein is dependent on ionic strength (Figure 3).

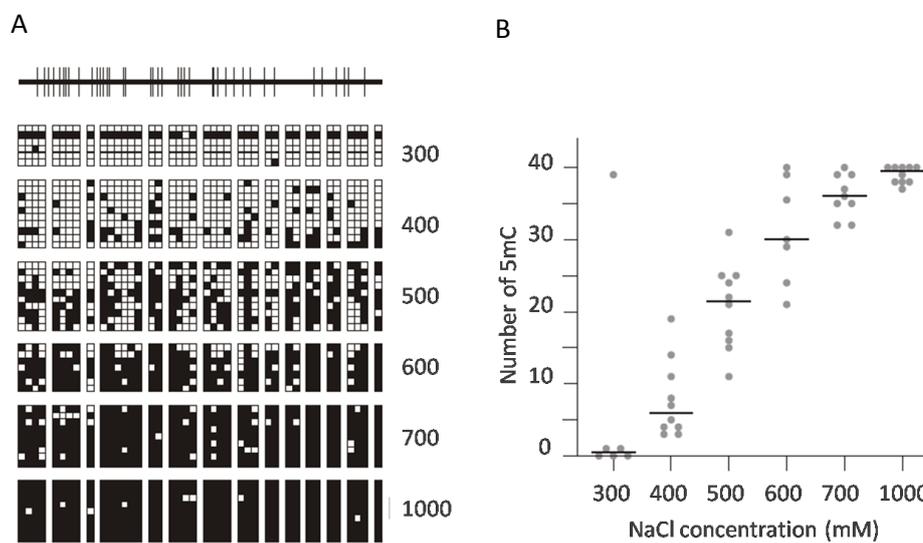


Figure 3: Enrichment of an *in vitro* partially methylated gene locus after MCIp.

A: Clonal bisulfite sequencing of DNA fragments fractionated according to increasing NaCl-concentrations in the buffer. The CGI depicted on top of the figure was bisulfite treated, amplified, cloned, and sequenced. Squares represent individual CpGs (empty: unmethylated; filled methylated). B: Scatter dot plot showing the numbers of methylated CpGs in a sequenced fragment according to the enrichment fraction. Horizontal bars represent median number of methylated CpG. Modified from [81].

Stepwise elution of the sequences from the MBD protein with increasing amounts of salt (300-1000 mM) results in enrichment fractions with increasing DNA methylation levels. By using the highest salt containing fraction (1000 mM) only or together with the lower salt-containing fractions (400-1000 mM), capture of high and medium-to-high methylated DNA fragments, respectively can be achieved [73]. For genome-wide analysis, microarray hybridization or NGS technology is used [83, 84]. MChIP-Seq does not provide precise quantification of methylation levels since differences between samples are measured as enrichment scores of fragments and the overall resolution is rather low (>100 bp). However, MChIP offers a way to reduce the complexity of the whole genome at moderate costs [85].

1.2.2.3 Reduced representation bisulfite sequencing (RRBS)

RRBS represents a quantitative bisulfite treatment-based method. It is a high-throughput technique to analyze methylation profiles using restriction endonucleases and bisulfite treatment [86]. With this technique single nucleotide resolution of quantitative methylation data is available on a genome-wide scale by requiring modest amounts of sequencing only [87].

The restriction endonuclease *MspI* is methylation independent and generates a number of DNA fragments including highly redundant fragments from microsatellite DNA. Fragments are predictable and constant, increasing the utility for comparative DNA methylation analysis in a given specimen. Digestion with *MspI* generates fragments with two terminal CpGs. Thus, each fragment provides at least one analyzable CpG measurement per end-sequencing read. Therefore, the number of reads required for high coverage and reproducibility of a certain CpG methylation value is reduced.

Small fragments sequenced by RRBS cover approximately 1 % of the total genome and 10 % of all available CpGs. Due to the *MspI* recognition site, genomic regions with a high CpG density are preferentially subjected to library preparation, resulting in a coverage of 75-90 % of all promoters and CGIs [74]. Restriction enzyme combinations can further increase the coverage for regions with lower CpG density such as CGI shores, intronic, and enhancer regions [88]

Several modifications have been reported to refine or adapt the RRBS procedure for multiplexing [89], for the use of FFPE material [90], for laser capture microdissected specimen [91], or for single-cell analysis [92].

1.2.2.4 EpiTYPER MassARRAY

Quantification of DNA methylation by Matrix-Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) Mass spectrometry is an alternative approach to sequencing- and hybridization-based methods (Figure 4).

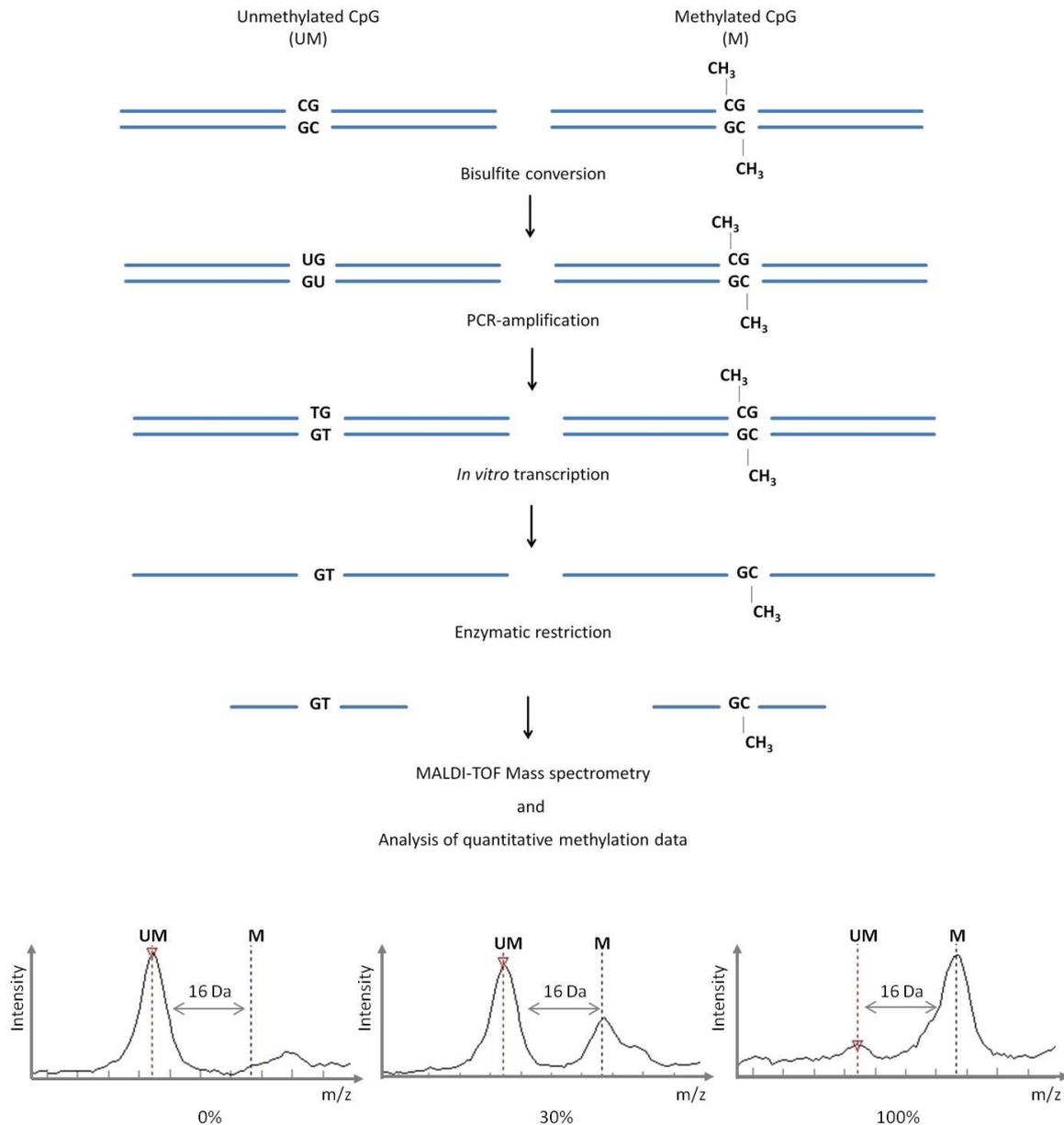


Figure 4: Schematic overview of the quantitative EpiTYPER technology.

During bisulfite conversion unmethylated cytosines (C) are chemically deaminated to uracils (U), while methylated cytosines are protected. In the subsequent PCR amplification thymine is incorporated instead of uracil translating the methylation mark into a genetic mark. The reverse strand is *in vitro* transcribed by T7 R&DNA polymerase and enzymatic restriction yields to short fragments that are analyzed using MALDI-TOF Mass spectrometry. Fragments are resolved by mass and charge as there is 16 Da mass shift between previous unmethylated (UM) and methylated (M) CpGs due to the change in genetic sequence. Three different methylation values (0 %, 30 % and 100 %) are depicted in the spectra, exemplarily (modified from [93]).

This technique is not genome-wide but is able to provide accurate quantitative methylation data after region-specific amplification. Since this method is implemented in the EpiTYPER platform from Sequenom, it is suitable for high throughput and sample automation [93, 94]. Primers are designed using the EpiDesigner software, optimal amplicon length vary between 150-500 bp depending on the quality of DNA available, *e.g.*, using FFPE material shorter amplicons are recommended.

Using bisulfite conversion the methylation information is translated into a genetic mark and stably transmitted through PCR amplification. The reverse strand is *in vitro* transcribed by T7 R&DNA polymerase. Subsequent base-specific cleavage yields to short fragments that are quantitatively analyzed using MALDI-TOF. Due to the shift in genetic sequence a 16 Da mass difference is introduced between unmethylated and methylated CpGs. Fragments are resolved according to their mass-to-charge ratio and are quantitatively analyzed by fragment peak calling.

This method is of particular value for technical validation/confirmation of genome-wide data and subsequent high throughput screenings in additional cohorts.

1.2.3 Histone tail modifications

The DNA in the nucleus is organized as chromatin, a DNA histone complex with repeating units of nucleosomes. 146 base pairs of DNA are wrapped around a histone octamer composed of two molecules H2A, H2B, H3 and H4 with H1 serving as linker. Histones are globular basic proteins with a flexible N-terminus ("tail") that can be subjected to a number of covalent but reversible post-translational modifications including acetylation, methylation, phosphorylation, ADP-ribosylation, ubiquitination and SUMOylation. Depending on the modified amino acid residue, histone tail modifications relax or compact chromatin structure and thereby, facilitate or suppress transcription of the downstream genes [95].

Histone acetylation is established by the action of histone acetyltransferases (HATs) and represents an activating mark opening up the chromatin structure. Thus, proteins with HAT catalytic activity such as p300/CBP (CREB-binding protein) or PCAF (p300/CBP-associated factor) are often transcriptional co-activators. Removal of histone acetyl groups by catalyzing the transfer to CoA is realized by Histone deacetylases (HDACs 1–11) which results in chromatin condensation and transcriptional repression.

Histone methylation is catalyzed by HMTs and takes place at lysine residues. So far, more than 50 HMTs have been identified in humans that transfer a methyl group from SAM to lysine residues [96]. Depending on the methylation status (mono-, di-, or trimethylation) and the lysine residue that is methylated, the histone mark possess activating (*e.g.*, K4, K36, K79 in H3) or repressive effects (*e.g.*, K9, K27 in H3) on gene expression. Trimethylation at H3K4 (H3K4me3) is mainly associated with active TSS, whereas H3K27me3 is a transcription repressing modification introduced by the multiprotein repressive complex PRC2 (polycomb repressive complex 2). Also H3K9me2/me3 exhibit suppressive effects on active gene transcription and is predominantly found in heterochromatin.

Histone lysine demethylases (HDMs) remove histone methylation marks, *i.e.*, by lysine specific demethylase 1 (LSD1) or the family of about 20 Jumonji domain-containing (JmjC) HDMs [95, 97].

During tumorigenesis HDACs and HMTs are often increasingly expressed leading to a dysregulation of repressive histone marks. Particularly EZH2 (enhancer of zeste 2), a member of the PRC2, has been found to be overexpressed in breast and prostate cancer [98]. HDMs are able to erase methylation residues from activating (H3K4me3) as well as from repressive marks (K3K9me3, H3K27me3) and have been reported to be upregulated in prostate cancer, although their context specific action has not yet been fully understood [99, 100].

1.2.4 Non coding RNA (ncRNA)

The area of epigenetic regulation by long non-coding RNAs (lncRNAs) is an emerging field in current cancer research. Some lncRNAs are long known, such as Xist for X-chromosome inactivation in mammals or H19 for genomic imprinting. Others have been identified recently such as HOTAIR with increased expression in primary breast tumors and metastases, SCHLAP1 as a marker and regulator in prostate cancer [101, 102], or overexpression of MALAT1 in non-small cell lung cancer [103].

The most well studied small ncRNAs are micro RNAs (miRNA or miRs); 18–25 nucleotide long single stranded ncRNAs which regulate gene expression at the post-transcriptional levels by silencing their target genes. As part of the “RNA-induced silencing-complex” (RISC), miRNAs sequence-specifically bind to the 3' untranslated region (3' UTR) of mRNAs and induce target degradation, or block translation in the case of perfect or partial base-pairing with the 3' UTR sequence, respectively [104].

Since 2002, when Calin et al. provided the first evidence that down-regulation or deletion of miRNAs plays a role in leukemogenesis a multitude of miRNAs have been identified as important regulators in carcinogenesis and development of other diseases [105]. About 2600 human, 2000 mouse, and 800 rat miRNAs have been identified, functioning as oncogenes (onco-miRs) or tumor suppressor miRs [106]. A number of miRNAs serve as diagnostic and prognostic tools since their expression profiles vary in tumor compared to normal tissue. Additionally, they have been suggested to classify tumors more efficiently than mRNA-based methods [107].

MiRNAs themselves can be affected by the regulation of epigenetic mechanisms, and, on the other hand, are able to target the 3' UTRs of writers, readers, editors and erasers of epigenetic marks thus modulating epigenetic regulation. The complex interactions of the epigenetic machinery, emphasizes their comprehensive probabilities that are available to maintain gene expression and cell identity.

Since alterations in gene expression *via* epigenetic mechanisms are reversible, the epigenome has been identified as an attractive target for chemoprevention with dietary factors. The impact of IF on DNA methylation of the breast will be introduced in detail in section 1.3.2.5.

1.3 Mammary glands, breast cancer and preventive impact of IF

1.3.1 Normal mammary gland development

The mammary gland (MG) is one of the few organs in mammals that develop after birth. Genetic and epigenetic mechanisms control mammogenesis characterized by dramatic changes in size, shape and function from lobule formation during puberty, development and differentiation during pregnancy and lactation, to regression after menopause [108].

Histologically, lobules and ducts are composed of a single layer of luminal epithelial and contractile myoepithelial cells originating from the same progenitor stem cells. They are oriented in a transverse fashion and surrounded by a basement membrane which separates them from the stroma. The surrounding tissue is composed of extra cellular matrix (ECM), fibroblasts, immune cells and adipocytes as well as blood vessels and comprise the local microenvironment (Figure 7A).

1.3.1.1 Childhood, puberty and adolescence

In newborns, there are primitive structures of ducts which just keep pace with the general growth during childhood. Main changes of the MG are initiated at puberty and can be defined from external appearance or by determination of the MG area, volume, and degree of branching, as well as the level of differentiation by characterizing the lobule type [108].

Ducts grow, divide and form terminal end buds (TEB, or terminal ductal lobular unit (TDLU)), highly proliferative structures which give rise to branches, ductules, and alveolar buds (AB), morphologically more differentiated structures. AB cluster around terminal ducts and form three different lobule types composed of increasing numbers of ductules with enhanced degree of differentiation and decreasing degree of proliferation (lobule type 1-3) (Figure 5).

Normal MG epithelium undergoes oscillating phases of cell proliferation and cell death during the menstrual cycle leading to phases of normal hyperplasia. Although this seems to be a balanced process mammary development never returns to the starting point of the previous cycle resulting in increased bud formation with every ovulatory cycle until age 35 [108].

1.3.1.2 Pregnancy, lactation and menopause

Maximum MG development is achieved during pregnancy by the control of prolactin and placental lactogenic hormones. The early stage pregnancy is characterized by proliferation of distal ducts, as they increase in size and number, resulting in acini formation and development of lobule type 4. In the later stage pregnancy, the secretory activity of the epithelium increases and marks the full term glandular differentiation. The proliferation of new acini is reduced and lipids get accumulated in epithelial vacuoles. Postpartum, prolactin induces lactation with colostrum being secreted during the first week, followed by secretion of transitional milk for 2-3 weeks and finally mature milk.

No major developmental changes are observed during the lactation period. The accumulation of milk in the acini after weaning has an inhibitory effect on milk production and lead to postlactational regression due to cell autolysis, collapse of acinar structures, phagocyte infiltration, and regeneration of the connective tissue. New proliferation and budding of the terminal ductules lead to more glandular tissue in the parous gland compared to the virgin gland until menopausal involution occurs.

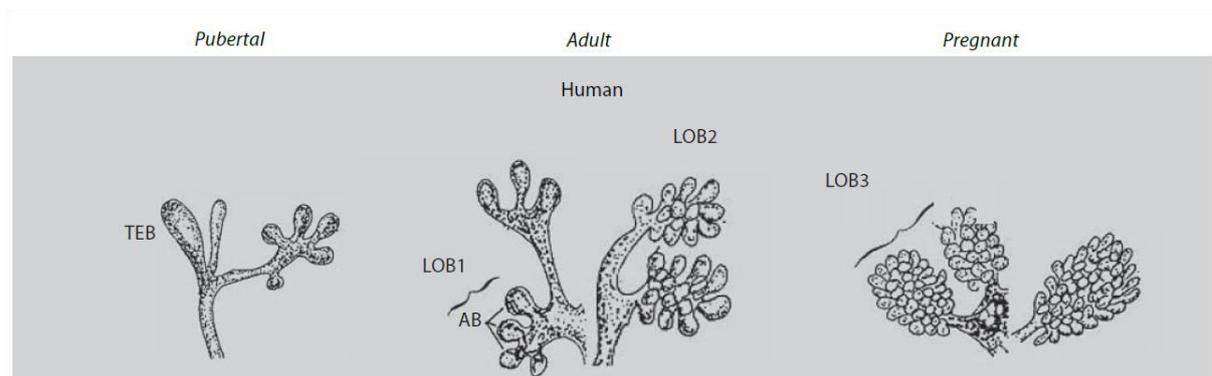


Figure 5: Developmental phases of the human MG.

TEB: terminal end bud, AB: alveolar bud, LOB1-3: lobule type 1-3 (from [109]).

During menopause, ovarian hormone production comes to an end leading to amenorrhea and regression of the MG to lobule type 1 of both, parous and nulliparous women. Although MGs might be composed of identical lobule types, the personal reproductive history might have introduced permanent changes in MG, potentially by epigenetic changes, and affect susceptibility of the breast to cancer development.

1.3.2 Breast cancer development

With an incidence of 1.67 M newly diagnosed women and 521,000 cancer death in 2012, breast cancer is the most common form of cancer affecting women worldwide [110-112]. In the US, it is

estimated that there will be 231,840 new breast cancer cases in 2015, and 40,290 women will die of the disease [113]. The development of breast cancer is highly dependent on endocrinological events and reproductive history. Hormone-related risk factors such as early onset of menarche (first menstrual cycle), late menopause and advanced age at first pregnancy lead to prolonged exposure to elevated serum levels of ovarian-associated hormones and thus increase the risk of developing breast cancer [114].

Whereas most breast cancers arise spontaneously, about 5-10 % might be inheritable, with roughly 30 % due to germline mutations in the susceptibility genes *Breast cancer 1/2, early onset* (BRCA1/2). Due to rapid expansion of massive parallel sequencing, additional genes that are involved in familial breast cancer risk have been identified, *i.e.*, TP53, PTEN, and ATM. Yet, it has not been clarified whether mutations in these genes are disease drivers or passenger effects [115].

1.3.2.1 Breast cancer progression

Human breast cancer development is a continuous but non-obligatory linear multistep process (Figure 6). There are several stages that defined breast cancer development from TDLU to invasive breast cancer (IBC). The shift from TDLU to premalignant hyperplastic breast lesions (hyperplastic enlarged lobular units (HELU)) is characterized by increased proliferation of the MG epithelium. Atypia (atypical ductal hyperplasia (ADH), atypical lobular hyperplasia (ALH)) differs from HELU by alterations in adhesion and cell polarity, since cells pile up and acini enlarge in size. By further expansion of the cell volume carcinoma *in situ* arise (ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS)). Appearance of biological and histological diversity distinguishes well- from poor-differentiated DCIS displaying slow and fast proliferative characteristics (low and high grade). Further accumulation of epigenetic and genetic alterations lead to breakdown of the basement membrane and spreading into the local microenvironment, ultimately called invasive breast carcinoma (IBC) [116, 117].

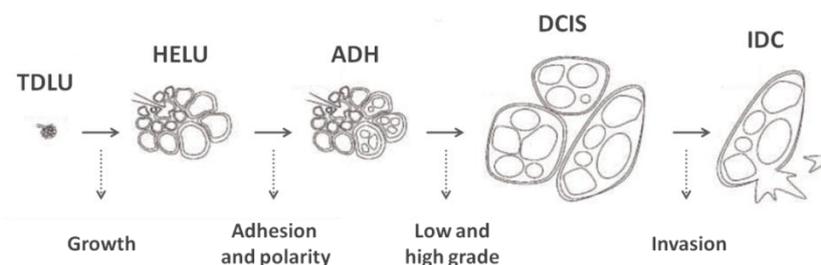


Figure 6: Model of human breast cancer progression focused on the evolution of ductal carcinomas.

TDLU: terminal ductal lobule unit, HELU: hyperplastic enlarged lobule unit, ADH: atypical ductal hyperplasia, DCIS: ductal carcinoma *in situ*, IDC: invasive ductal carcinoma (modified from [117]).

During this process, alterations in the ECM and neo-angiogenesis facilitate the migration of tumor-associated macrophages (TAM). Growth factor production by TAM and CAF (cancer-associated fibroblasts) influence tumor cell migration into the bloodstream. Circulating tumor cells (CTC) that escaped the primary tumor eventually form metastasis at distant sites (Figure 7B).

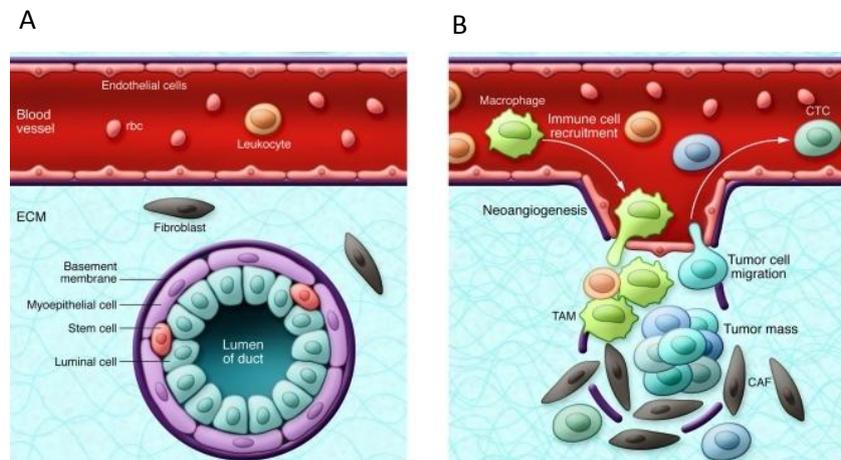


Figure 7: Schematic representation of the normal MG and invasive breast carcinoma.

A: Normal breast architecture and B: microenvironmental alterations in invasive breast carcinoma. ECM: extra cellular matrix, rbc: red blood cell, TAM: tumor-associated macrophage, CTC: circulating tumor cell, CAF: cancer associated fibroblast (modified from [118]).

1.3.2.2 Breast cancer classification

The term breast cancer includes a number of heterogeneous tumor entities with different clinico-pathological characteristics. Multiple subgroups of breast tumors can be classified based on classical histological and immunopathological evaluation, hormone receptor status, grading, and molecular signature [118]. Invasive ductal carcinoma (IDC, 75 % of cases) and invasive lobular carcinoma (ILC, 10 % of all cases) are the most frequent histological subtypes of breast cancer, accounting for roughly 90 % of all cases. The remainder adenocarcinomas are histologically special types, subcategorized as, *e.g.*, apocrine, comedo, medullary, mucinous, neuroendocrine, and many more [119, 120].

The presence of specific histological and immunopathological markers have been used to define tumors and to identify adequate treatment therapies. Clinically interesting markers such as ER and progesterone receptor (PR, encoded by Pgr) as well as the oncogene human epidermal receptor 2 (Her2, erb-b2 receptor tyrosine kinase 2 (ERBB2)) are used to assign specific categories, *i.e.*, ER positive, HER2 positive, triple negative (ER⁻, PR⁻, HER⁻), and triple positive tumors to predict the benefits of endocrine therapy using ER-antagonists and aromatase inhibitors, or monoclonal antibodies against HER2, respectively. Molecular subtypes add another layer of detail to the highly diverse ductal and lobular tumors by looking at distinct gene expression patterns. Five main subtypes

have been classified to predict patient relapse, overall survival, and response to treatment. Luminal A and B subtypes are both ER positive and display the best prognosis for patient survival. Normal breast-like cancers are associated with gene expression profiles similar to adipose tissue, and basal-like and ERBB2 positive (HER2 overexpressing) tumor subtypes have the least favorable outcome overall [120-122].

1.3.2.3 Epidemiology of breast cancer and the impact of IF

The incidence of certain hormone-dependent cancer types such as breast and prostate cancer has been found to be lower in Asian countries than in Western populations (age-standardized number of new cases per 100,000 inhabitants in 2012: breast cancer 29.1 vs. 96, prostate cancer 9.4 vs. 95) [111]. Furthermore, migration studies showed that Japanese and Chinese American women had a 60 % higher risk to develop breast cancer if they were born in Western countries compared to those born in the East. If their grandparents were also born in the West women possessed an additional 50 % higher breast cancer risk [123].

These studies indicate that lifestyle and environmental factors such as dietary habits, essentially affect international incidences of certain cancer types [124, 125]. Since hormone-dependent cancers rates are particularly low in populations following a traditional Asian diet rich in soy and soy products, IF have attracted much attention regarding their chemopreventive potential.

Results from epidemiological investigations about beneficial health effects of IF are actively discussed [126, 127]. The majority of the case-control studies have shown an inverse association for soy food intake, serum GEN levels and breast cancer risk, particularly in Asian women [128-131]. In addition, experimental investigations in rodents highlight the protective effects of GEN, especially for chemically induced cancer. In contrast, rodent data investigating the impact of GEN on the growth of already existing tumors and in xenograft models in immunodeficient mice, revealed stimulatory and growth promoting effects on human breast cancer cells [132]. In a further study of the same authors, differently processed soy products were investigated to affect growth of MCF-7 breast cancer cells transplanted into ovariectomized athymic mice, all providing the same amount of GEN equivalents in the diet (750 ppm) [133]. Interestingly, soy extracts (NovaSoy650®), mixed IF or GEN alone, but not soy flour increased tumor growth and mRNA expression of estrogen-responsive genes (*e.g.*, pS2, cyclinD1). These result demonstrate that IF administered within a complex food matrix show drastically different effects on cancer progression compared to isolated IF [134]. The reason for this might be the presence of other bioactive compounds like peptides, protease inhibitors, phytic acid and saponins and probably other phytoestrogens which differently affect gene expression [135-138]. However, a recent meta-analysis investigating 84,450 women in a multiethnic cohort did not find a

statistically significant association between IF intake (0-180 mg/d) and breast cancer risk. The authors suggested possible ethnic/racial differences in risk estimates and an inverse association in populations consuming high amounts of soy products [139].

The timing of exposure seems to be the most critical factor [140]. It has been shown that especially a multigenerational and prepubertal exposure to IF modulate MG morphology and provide anti-tumorigenic effects [141]. In a study of Lamartiniere et al., prepubertal IF exposure accelerated differentiation of the epithelial cells by transient upregulation of EGF-signaling and protected against chemically induced cancer in rats. Following this, EGF-receptor (EGFR) expression was persistently reduced in TEBs leading to an increased number of more differentiated alveolar buds [142]. In contrast, a short term high dose IF supplementation in pre- and postmenopausal women stimulated epithelial cell proliferation and resulted in increased mammographic density (representing a risk factor for developing breast cancer [143]).

Since many processes during early development are regulated by epigenetic mechanisms, it is tempting to speculate that early epigenetic reprogramming of the MG is targeted by IF, affecting normal cell growth and susceptibility to breast cancer [144].

1.3.2.4 DNA methylation changes associated with breast cancer

Several studies investigating tumor-associated candidate genes involved in DNA repair, apoptosis, or cell cycle control have shown breast cancer specific aberrant DNA methylation which arises early during disease progression, *i.e.*, of APC, RAR β , RASSF1A, HIN-1 and SFRP1 [145, 146]. With the availability of genome-wide methylation data of breast tissue, distinct patterns have been associated with luminal and myoepithelial differentiation [118]. Luminal genomes harbor twice as many hypomethylated enhancer elements and four times the transcriptional output (RNA yield/cell) when compared to myoepithelial cells [147]. This transcriptional amplification in luminal cells might be of advantage in disease initiation and progression since this phenomenon has been described for c-Myc driven tumor types. Less transcriptionally active myoepithelial cells are commonly absent in breast tumors [148].

Moreover, distinct DNA methylation patterns have been associated with molecular subtypes [149]. Particularly, luminal breast cancers are characterized by a higher frequency of DNA methylation gain compared to basal-like tumors. Further studies reported associations with hormone receptor and mutational status as well as disease progression and prognosis [150]. For a comprehensive review see [151].

1.3.2.5 Impact of IF on DNA methylation in breast cancer

Early studies on DNA methylation were mainly interested whether soy IF are capable in reactivating TSGs silenced by DNA hypermethylation and were mostly performed in cultured breast and prostate cancer cells. Particularly, genes were investigated that were involved in sustained proliferation and resisting cell death, *i.e.*, RAR β 2, RASSF1, PTEN, cyclinD2 and in DNA repair, *i.e.*, BRCA1, BRCA2 as well as detoxification, *i.e.*, GSTP1 and revealed IF-induced demethylation of the promoter region [152-155]. As an underlying demethylating mechanism IF have been shown to inhibit DNMTs and MeCP activity *in vitro* which might result from enzyme inhibitory effects or down regulation of gene and protein expression [156].

Treatment of breast cancer cell for 1-2 weeks with low GEN concentrations (3.125 μ M) slightly reduced methylation as measured by methylation-specific PCR of the promoter region of GSTP1, a Phase II metabolizing enzyme commonly silenced in breast and prostate cancer, with subsequent but weak mRNA re-expression in ER α -negative MDA-MB-468 but not in ER α -positive MCF-7 cells [152]. The promoters of two TSGs RAR β 2 (Retinoid acid receptor β 2, a nuclear transcriptional regulator involved in cell growth and differentiation) and HIN1 (high in normal-1, aberrant methylation serves as biomarker for prognosis in breast cancer) could be demethylated using the same dose of GEN in another ER α -positive cell line MCF-10a.

In T24 Ha-ras transformed MCF-10aT and MCF-7 cells Li et al. investigated the effects of higher GEN doses on expression of the catalytic subunit of human telomerase (hTert, human telomerase reverse transcriptase) [157]. hTert is involved in telomeric stabilization and is normally repressed in postnatal somatic cells resulting in progressive shortening of telomeres as an important component of cellular aging. Reactivation of hTert is a crucial event during cell transformation and promotes proliferation and survival potential. Treatment with high doses of GEN (50-100 μ M) increased expression of E2F1, a repressor of hTert transcription, resulting in enhanced binding to the hTert core promoter and reduced levels of hTert mRNA expression. GEN also reduced protein expression of DNMT 1, 3a, and 3b, and bisulfite sequencing data indicated a shift from methylated to unmethylated CpGs in the E2F-1 binding site. Although the impact of GEN on methylation was rather weak, the results suggest that GEN is working as an epigenetic modulator and might facilitate preventive approaches in breast cancer.

Bosviel et al. showed the impact of moderate doses of GEN (18.5 μ M), DAI 78.5 μ M) and Equol (2 μ M) on BRCA1 and BRCA2 involved in DNA repair mechanisms and often mutated in breast cancer. Changes in methylation levels were generally low but IF treatments further demethylated CpG sites

in the promoter and/or exon region, resulting in a weak increase of gene expression in three breast cancer cell lines [153].

High concentrations of GEN (60, 100 μ M) dose- and time-dependently reduced DNMT1 mRNA and protein expression as well as DNMT activity in MCF-7 and MDA-MB-231 breast cancer cells [156]. These observations were accompanied by a decrease in cell viability and a global loss of DNA methylation. Specifically, promoters of four TSGs that are often deregulated in familial breast cancer including ATM (ataxia telangiectasia mutated, an important cell cycle checkpoint kinase), APC (adenomatous polyposis coli, an antagonist of the Wnt-signaling pathway), PTEN (phosphatase and tensin homolog, a negative regulator of AKT/PKB signaling), and SERPINB5 (mammary serine protease inhibitor encoding the anti-proliferative Maspin protein) were hypomethylated due to the GEN treatment, resulting in increased gene expression levels.

So far, there is only limited information on the impact of soy IF on DNA methylation in breast tissue *in vivo*. Li et al. investigated the effects of IF on ER-negative breast cancers. Treatment of ER α -negative MDA-MB-231 cells with GEN induced re-expression of ER α and ER β through increased histone acetylation at the ER α promoter. Additionally, a re-sensitizing for E2 and tamoxifen treatment was demonstrated, especially in combination with Trichostatin A, a HDAC inhibitor, demonstrating an interesting strategy for clinical implications. Accordingly, dietary GEN (250 mg/kg diet) inhibited growth of breast cancer in a MDA-MB-231 xenograft model and increased therapeutic sensitivity of tamoxifen [158].

In a prospective, randomized, double-blind intervention trial, 34 healthy premenopausal American women received capsules with soy IF representing a typical Asian diet through one menstrual cycle (low dose: 37 mg/d; high dose: 128 mg/d, 70 % GEN, 27 %DAI and 1.3 % GLY) [159]. Serum GEN levels dose-dependently increased after the intervention leading to lower complement C3 levels as a measure of anti-estrogenic activity of the IF intervention (nonsignificant trend). DNA methylation was analyzed in breast tissue obtained by mammary ductoscopy. Five candidate genes were investigated including p16, RASSF1, RAR β 2, cyclinD2, and ER all frequently silenced by methylation in breast and prostate cancer and counteracting sustained proliferation potential as well as resisting cell death. No significant change in DNA methylation could be detected. However, cyclinD2 and RAR β 2 promoter methylation increased with high and decreased with low GEN serum levels, respectively, suggesting a different mechanism of action for high vs. low GEN levels in healthy premenopausal women.

Of note, methodologies used in these studies often did not provide quantitative methylation information (*e.g.*, methylation specific PCR) and mainly focused on selected candidate genes.

Genome-wide studies investigating the impact of IF on the methylome would permit a more comprehensive evaluation of the IF-induced effects on the reprogramming of the MG and consequently on cancer prevention.

1.4 IsoCross project

The collaborative research project IsoCross: Isoflavones: Cross-species' comparison on metabolism, estrogen sensitivity, epigenetics and carcinogenesis is a DFG-funded research initiative. It aims to identify molecular mechanisms which determine whether the exposure towards IF may be adverse or protective regarding the risk to develop E2-induced breast cancer, taking into account the influence of dietary fat consumption and obesity. The main research concept is centered on IF metabolism and bioavailability, E2 levels and metabolism, modulation of tissue selective E2 sensitivity, and carcinogenesis of the breast. The strength of the IsoCross project is the integrative analysis of metabolite levels and molecular mechanisms including epigenetic modifications as well as the cross-species' comparative interpretation of the results of animal experiments and human intervention studies.

The research project was a collaborative effort of an interdisciplinary team originating from five scientific institutions in Germany. Our collaboration partners Prof. Diel from the German Sports University Cologne (DSHS) and Prof. Vollmer from Technical University Dresden (TU) performed animal experiments in rats, addressing specific research questions on dose and timing of soy exposure, influence of dietary and body fat, and influence of soy on E2-induced carcinogenesis, providing us with MG tissue for epigenetic analyses. Profs. Kulling and Bub from the Max Rubner Institute Karlsruhe (MRI) and Prof. Lehmann from the University Würzburg focused on the human aspect of the cross-species' comparison. Both groups performed human intervention studies with soy extracts. Their main interest was on IF and E2 metabolism and levels of IF and E2 metabolites in plasma, but also in target tissue such as MGs. Human tissue material including PBMCs, fat biopsies and breast tissue are currently investigated in bachelor and master thesis projects.

The initial focus of the present Ph.D. project was on the impact of IF on DNA methylation during E2-induced breast carcinogenesis in ACI rats. Since the establishment of the model took more time than initially anticipated due to problems with E2-releasing pellets, we took the opportunity to address additional aspects of potential IF impact on epigenetic mechanism during MG development, making use of tissue material from the other sub-projects in the IsoCross project. The aims of the thesis reflect the multitude of addressed research questions.

2. Aims

There are several investigations that provide evidence for the impact of IF on the epigenome in cancer prevention. Particularly, research has been done with cancer cell lines and focused on selected candidate genes. So far, studies investigating genome-wide changes on DNA methylation upon IF treatment are scarce. Also, previous work was mainly carried out in cancer cell lines which might have acquired additional methylation changes during *in vitro* culture that might not be susceptible to alterations by intervention with dietary factors. In addition, the activity of IF to sustainably affect DNA methylation patterns might be more pronounced when analyzing effects at specific time windows of susceptibility and active reprogramming during development.

In order to address some of the limitations of the existing studies with IF on DNA methylation, the aim of this thesis was to analyze DNA methylation profiles in MGs of healthy rats and during E2-induced mammary carcinogenesis in rats on a genome-wide scale.

1. In the first part, the goal was to clarify the impact of soy IF at different doses and critical time windows on DNA methylation in MGs in healthy Wistar rats. In particular, it was of interest:
 - whether a short term high dose IF exposure after menopause (ovariectomy) as in Western countries induces a contraindicative pro-proliferative effect on the MG epithelium?
 - whether dietary IF exposure during puberty, representing the major postnatal phase of MG development besides pregnancy, modulates the MG methylome sufficiently as seen for lifelong IF exposure?
 - whether low dietary IF doses are sufficient to induce potential beneficial health effects?
2. In the second part, ACI rats were lifelong exposed to high doses IF to assess the influence on DNA methylation during E2-driven carcinogenesis of the MG. The following questions were addressed:
 - does dietary IF exposure during developmentally critical time windows, *e.g.*, embryonic/fetal, early postnatal and pubertal periods as in Asian countries, modulate the epigenetic programming of the MG?
 - consequently, does this potential impact on DNA methylation lower the susceptibility to breast cancer and reveal cancer preventive effects?

Reports on genome-wide methylation profiles of rat tissues are very limited and reference methylomes are not available. Therefore, the project included the establishment of two sequencing-based methodologies for methylation profiling, *i.e.*, MCIp-Seq and RRBS as well as bioinformatics evaluation of results.

3. Material and Methods

3.1 Material

3.1.1 Instrumentation

Equipment which is not standard lab equipment is listed here or mentioned in the methods section.

Name	Manufacturer
MassARRAY Compact System	Sequenom, Hamburg, Germany
MassARRAY Nanodispenser	Sequenom, Hamburg, Germany
Mastercycler eppgradient384	Eppendorf AG, Hamburg, Germany
Mastercycler eppgradient	Eppendorf AG, Hamburg, Germany

3.1.2 Software and databases

Name	URL
Gene Set Enrichment Analysis	http://www.broadinstitute.org/gsea/index.jsp
Adobe Illustrator	https://www.adobe.com/
EpiDesigner	http://epidesigner.com/
EpiTYPER	http://www.massarray-epityper.software.informer.com/
EulerAPE	http://www.eulerdiagrams.org/eulerAPE/
Gene Expression Omnibus	https://www.ncbi.nlm.nih.gov/geo/
GraphPad Prism®	http://www.graphpad.com/
Ingenuity Pathway analysis	http://www.ingenuity.com/products/ipa
IGV 2.1	http://www.broadinstitute.org/igv/home
Homer Suite of tools	http://homer.salk.edu/homer/
Multi Experiment Viewer	http://www.tm4.org/mev.html
Primer3Plus	http://www.primer3plus.com/
Qlucore	http://www.qlucore.com/
R	http://www.r-project.org/
RnBeads	http://rnbeads.mpi-inf.mpg.de/
SNPs info rat	ftp://ftp.ncbi.nih.gov/snp/organisms/rat_10116/
UCSC Genome Browser	http://genome.ucsc.edu/

3.1.3 Primers

3.1.3.1 MassARRAY primers

Gene Name	Left	Right	Next TSS [bp]	Genomic Annotation	Length	CpGs	T _M [°C]
Aldh1L1	TTGGTTATATTTTGATGTGATGGAG	AAACCATTAATAAAACAACAATC	-664	promoter-TSS	423	18	58
Arpp21	TTTGGGAAAAGAAATTGGTATAAA	AAAAACACATTAATTACAAAACCA	-1247	promoter-TSS	392	8	60
Bpgm	GGGATTTGGTATTAGAGTTGGAAGT	TCTTCCTAAATATATAAAAACTAAACCCA	822	promoter-TSS	265	5	60
Cdh13	ATAGTAAGAATTTTGGTTGGTTGGA	AACTCTAAATACCACACCCATTAC	304332	intron (2 of 13)	336	6	58
Cldn4	TTTGGAGAATGTGAGGGGTTATAG	CAAAACCAAATCATAATCACAACC	716	promoter-TSS	376	20	58
Cyp2C6v1	TTGATTTTTGTATTTAATAGGATGG	AAACAACAAAACCAAATTAATCC	9598	exon (3 of 9)	324	5	58
Esr1	TTTTTAGAAATGTTTTATGGTTTTG	TATCCCTCTCTTTAACAAAAACAC	-72	promoter-TSS	379	14	60
Extl1	GTTTTTAGTATGTTTTGGATATGGT	AAAACAAAACACAACCATATACCC	-1851	promoter-TSS	394	12	60
Fxyd2	GGATTTATAAGTTAGTTTTTAGTATTTA	AACAAATATTCAACCTTAATTCCTTCA	-957	promoter-TSS	94	4	60
Gadd45b	GGAAGATTTTTGTGTGTTTTATGG	AAACCCTTACACTTTCCTTTCATA	6141	intergenic	328	19	58
Gbp2	GGGGGTAAGTTTTTAGTTGTTTTT	CATACCAAAATATCTCCCCACTAA	-2504	5' UTR	276	7	60
Gsn	GTTAGTGGGAAGGGAATTTAGATGT	ACAACAAAACATACTTACAAAACAAC	59	promoter-TSS	452	32	60
Lect2	TTTAGTTAGTAGTTGTAGGAAATTTGTGGT	TCCTAAAACAACCTATCAACCTAACAC	983	promoter-TSS	224	4	60
miR-10a	GAGAAGGGTAAGTTGAATTAATTTTTG	AAAAAACCAAAACCCCTC	-2532	intergenic	324	16	60
Mmaa	TTTTTGTGTGGTTTTTAAGGTTG	AAATCCAATTACAAACAAATAACTATTCC	-14004	promoter-TSS (splice variant)	240	12	60
Mrc2	TTGTTTTGTGTTGTTGGAATTTAGA	AATATCCCAAAAAACCTACATCATTC	-1038	promoter-TSS	267	8	60

Msx1	TAGGTTTTGTGGGTAGAGGTGAG	TCTCAAAAAACACAAAAACTACC	-1530	promoter-TSS	234	12	60
Ndr3	TTGTAAGTTTGGGTTTTATTTTAGTA	ACCTCAAACACTCAAACACACAT	1578	promoter-TSS	352	4	60
Niacr1	AAGTGGTTTTGTTGTTTATGAGGA	AACAACAAATTACACAAAACTCC	-98	promoter-TSS	235	7	60
Npas1	TTTTTTGGGGTTTTAGGATAATAGA	TCACAAATAAAAAACCATACACACA	3492	intergenic	356	11	60
Nr1h3	GTTGGGAATATAGGTTGGGGTTTA	AAACTAACTCCTCCCAAACAAC	1335	promoter-TSS	297	13	60
PTEN	TTGTTGAGGTTTGGATAGATTGA	ATCTCAAACAACTAACAAAAATA	61027	intron (7 of 8)	460	10	58
Rbm42	TTGGTTGGAGTTTAGGTATGAAAT	ATAATCCCAACTTAAAACTAC	-524	promoter-TSS	229	7	60
Rita	TGTTGGGTTGTTAAATAGGGTTTTA	TAACTCACTCTAACCTCAAAAAA	-851	promoter-TSS	206	7	60
Rnf8	GGTTTTAGGTGTTGGTATAGGTGG	TCCCTACAACTCAAAAAACTCA	2202	intron (1 of 7)	287	7	60
Sall3	GTTTATTAGGGTATTGGTTG	TAACACAATTCTCACAAAATAC	151	promoter-TSS	169	7	60
Shroom2	TGTTTGGTATGTATGGTGTTTAGA	AAAATAACTACAAACCAACCCCTTC	-714	promoter-TSS	500	12	60
Sirt4	TGAGTATTGTATTAGGGGTATGAA	AACTAACCTCAATTCTCCTCTCACC	2223	intron (2 of 4)	404	9	60
St3gal4	GGAAGGTTAAGAGTAGGTTTGT	CAAACAATTACATTACAATAACAACATCA	138471	intron (4 of 15)	432	13	60
Thy1	GTGTTGGTTTTTTGAGTTGTTGTTA	TTAAACCCCATCAAATTTAAACCTT	-113190	intergenic	437	8	60
Tlr3	TATGGTTAGAGGTTGGGATTGTAA	CCCCATTTATTTCTAAAAATCCCT	258	promoter-TSS	271	7	60
Wap	TGTGAGTTTTGTTTTGGTAGTTGG	AAAAATATCCTTTTATTCCTTATCTCC	3	promoter-TSS	385	7	58
Wif1	GAAAGGGGAAGATAATTTTTAGAGT	AAATCCTAACCTCTAAAACACTCC	38674	intron (2 of 9)	436	10	58
Zfp507	AAATTTTTAGTTGGGGTTATAGAGA	TACAACCTCTCCCTAATACTACC	-1902	promoter-TSS	400	10	60

Additionally added to left primer: 10 mer tag (AGGAAGAGAG) and to right primer: T7 R&DNA polymerase primer tag (CAGTAATACGACTCACTATAGGGAGAAGGCT).

3.1.3.2 Primers for genome-wide methylation profiling

Method	Name	Primer sequence	T _M [°C]
MCIp	Mest Left	GCCATCCCCGATCCTGTACC	60
	Mest Right	GGCCGCATTATCCCATGCC	60
	MCct Left1 (Spike in)	GGTTCTCTGCTGCCTTGC	60
	MCct Right1 (Spike in)	ACCCTCACCTCAACACCAC	60
	MCct Left2 (Spike in)	GCCCATGAGGCCAGTTAAT	60
	MCct Right2 (Spike in)	AGGCGCGCCATGAGCTC	60
RRBS	PPC Left	AATGATACGGCGACCACCGAGATCTACAC	60
	PPC Right	CAAGCAGAAGACGGCATAACGAGAT	60
	TruSeq Universal Adapter	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCT	
	TruSeq Indexed Adapter	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-nnnnnn-ATCTCGTATGCCGTCTTCTGCTTG	

3.1.3.3 Roche UPL qPCR primers

Name	UPL Probe	Left	Right	T _M [°C]
Aldh1L1	20	CATGTCCATCCAGACCTTCC	GTGGCACCCCTGGATCTTATC	60
Arpp21	1	CCTATAAGAAAAGACAGCTCTTTCG	TTGGAGCTGTCAGAATCTGTG	60
C3	116	TCATCCAGACAGACAAGACCA	TCCACAGTGAAGATCCGATAGA	60
Cdh1	13	CATGCTCTCTGTGGTCAAGC	TCATTCTCCACTTTGATCAGCA	60
DNMT1	16	AACGGTGTTGTCTACCGACTG	GCTGGCCATTTTGATGTTG	60
DNMT3a	1	CAGAAGTGCCGAAACATCG	GCACATCCACCAATGAAGA	60
DNMT3b	21	GCGATGGCAAGTTTTCTGAG	GGTGGCCAGATTAAGTGCT	60
Esr1	120	CCCTCCCGCCTTCTACAG	TCCTTGGCAGACTCCATGAT	60
Esr2	3	TCTGGGTGATTGCGAAGAGT	TTCCATGCCCTTGTTACTGA	60
Extl1	111	GCAGGTGGCGTATTACAACC	TGGGAAGTATCTCTCTGGGTAG	60
Foxa1	114	GAAAGGCTAGCCAGCTAGAGG	GATGAAGAAGCTGGAGACTTCAA	60
Fxyd2	73	CGGAGAATCCCTTCGAGTATG	TCTTACTGCCCCACAGC	60
Gadd45b	101	CGGCCAAACTGATGAATGT	TTGGATCAGGGTGAAGTGAA	60
Gata3	66	CTATGTGCCGAGTACAGCTC	ACATCCGAACCCGGTAGG	60
Greb1	120	CGTGCTATCAAAGGGACCTC	GGCAAAGAGCAGAGGATTGT	60
Gsn	77	CCCAAAGTCGGGTGTCTG	CTTCCCTGCCTCAGGAAT	60
Hprt	22	GGTCCATTCTATGACTGTAGATTTT	CAATCAAGACGTTCTTTCCAGTT	60
Hprt	95	GACCGGTTCTGTCATGTCG	ACCTGGTTCATCATCACTAATCAC	60
Krt8	67	CAACAAGTTCGCCTCTTTCA	GTTGCAGCAAGCTCCATTT	60

Krt5	120	CCCTGGATGTGGAGATCG	ACTCAGCCTGCACTCCTGTC	60
Ki67	110	AGATTCAGATTGTCCAGAAGTAAGC	TCAGACTCAAACCTCCACAAAGG	60
Mrc2	46	CAGGGCACTGCAGAGGAG	AGGATGGGCTATGCCATGT	60
Ndr3	62	GTGTTCTACTTTACTGGTGGTTGG	TAGGGTCGAGGCGAGAATTA	60
Niacr1	82	CAAACACAACGTGAGGAACCTG	GTGGTGAAGAAGGCCAAGTC	60
PCNA	41	ACCTCACCAGCATGTCCAA	CATAGTCTGAACTTTCTCTTGATTG	60
Pgr	20	CGCCCTACCTCAACTACCTG	GGTAAGGAATCAAAGCCATATTGT	60
PTEN	109	CGATACTTCTCTCAAATTTTAAGGT	TCTATAATGATCAGGTTTCATTGTCCT	60
Rnf8	101	AGCAGCTGGAGAAGACGTTT	TCCTTGCTCTTTCTCAAACC	60
SMA	25	CCCAGCACCATGAAGATCA	CGCCGATCCAGACAGAAT	60
β -actin	64	CCAACCGCGAGAAGATGA	CCAGAGGCGTACAGGGATAG	60
ST3gal4	128	CATGACCAGCAAATCTCACTG	TCTCGGAGACGGGAAAATAA	60
TBP	129	CCCACCAGCAGTTCAGTAGC	CAATTCTGGGTTTGATCATTCTG	60
Tet1	67	CGCCAGTTGCTTATCAAATC	GCCCCTTTCATTACCAAGTC	60
Thy1	40	CCACAAGCTCCAATAAACTATCAA	AGCAGCCAGGAAGTGTTTTG	60
TP63	74	CAGCACACGATCGAGACG	AGACTGAGACTGCATCGAGGT	60
Wap	120	CCACACCCAAATTCTGAGC	CCACAGGAACGACACCAGA	60
Wif1	68	GGCTCTGGAGCATCTACCT	CCCTCCGAGACAATCAGAAT	60
Zfp507	108	CCAATGTAAGCAGTGCAAAGAA	AAGGTGTTGGGCATGCAG	60

3.2 Methods

3.2.1 Experimental diets, animal subjects and study design

In order to investigate the impact of IF on DNA methylation *in vivo* two intervention studies were performed by our cooperation partners in the IsoCross Project as described below. In the first study, Wistar rats were exposed to dietary IF at different doses for various time periods, while the second study focused on the impact of a lifelong dietary IF on the kinetics of DNA methylation changes in MGs and on cancer preventive efficacy in the ACI rat model of E2-induced mammary carcinogenesis.

3.2.2 Soy isoflavone enriched diets

The studies were performed using a commercially available soy bean extract NovaSoy650[®] which was characterized in detail at the Max Rubner Institute in Karlsruhe by the group of Prof. Dr. Sabine Kulling (for details see [160]) (Table 1).

Table 1: Diets and soy extract applied in the Wistar and ACI rat experiments.

Name	Manufacturer
NovaSoy650 [®]	ADM, Decatur, IL, USA
R/M-H low-phytoestrogen complete-feed	Ssniff GmbH, Soest, Germany
SM R-Z protein-reduced complete feed	Ssniff GmbH, Soest, Germany

The diets investigated comprised (summarized in Table 2):

- 1) an IF-depleted diet (control (C)) purchased from Ssniff (R/M-H low-phytoestrogen complete-feed) which was free of any soy or alfalfa as protein or fat sources and contained approximately 3 ppm of IF aglycone equivalents.
- 2) a standard IF-containing diet (IF-medium (M) purchased from Ssniff (SM R-Z protein-reduced complete feed) which contained soybean meal and full fat soybeans as protein and fat source, resulting in approximately 232 ppm of IF aglycone equivalents.
- 3) specific IF-enriched diets. IF-low (L) and IF-high (H) diets were manufactured by adding NovaSoy650[®] at the amount of 0.131 g/kg (L) or 1.05 g/kg (H) to the control diet. These diets contained approximately 68 and 503 ppm of IF aglycone equivalents, respectively.

Table 2: Diet groups investigated in Wistar and ACI rat experiments.

Abbreviation	Diet name	IF aglycone equivalents [ppm]	NovaSoy650® [g/kg]	Ssniff complete feed
C	Control	3	-	R/M-H low-phytoestrogen
L	Low	68	0.131	R/M-H low-phytoestrogen
M	Medium	232	-	SM R-Z protein-reduced
H	High	503	1.05	R/M-H low-phytoestrogen

3.2.3 Rat uterotrophy assay in Wistar rats

Animal experiments were performed with Wistar rats at the German Sports University in Cologne by the group of Prof. Dr. Dr. Patrick Diel. For a summary of experimental outlines see Figure 8.

Adult female Wistar rats were obtained from Janvier Laboratories and were used for experimentation after two weeks of acclimatization (Table 3). These rats were maintained by Janvier Laboratories on a standard SM R-Z protein-reduced IF-containing diet (M) and were exposed to medium levels of IF. On PND 80 all females were ovariectomized and randomly split into two treatment groups. At PND 94 after 14 days of endogenous hormonal decline one group of the animals was subcutaneously treated for three consecutive days with E2 (4 µg/kg bw/day). Detailed information on animal handling, experimental conditions and E2-treatment is published in [160]. In order to mimic the postmenopausal situation with a short but high dose IF exposure scenario, rats were subjected to a diet switch to IF-high diet (MH group) or receiving control diet (MC group) during the hormonal decline period from PND 80 until PND 97. In order to control for effects due to the medium lifelong IF exposure until OVX in this intervention scenario, a second group of Wistar rat were subjected to the same diet swap protocol after ovariectomy (OVX) and hormonal decline until PND 97. Here dams and offspring were maintained on the control C diet and were not exposed to any IF prior to the post OVX IF-high diet (CH group).

Table 3: Experimental animals.

Name	Manufacturer
Wistar rats	Janvier Laboratories, Le Genest-St-Isle, France
August Copenhagen Irish (ACI/segHsd) rats	Harlan, Eystrup, Germany

For a lifelong IF intervention protocol, Wistar rats were obtained from Janvier Laboratories and were randomly assigned to control C or specific IF-enriched L and H diets prior to mating. Dams were kept

on allocated diets, thus female offspring were exposed to IF during conception, the fetal *in utero* period, as well as through weaning, and adolescence until the end of the experimental period in adulthood after OVX (PND 97, groups LL and HH).

The pubertal period as the major period of postnatal MG development beside pregnancy is a critical time window for epigenetic reprogramming [144]. Pregnant dams were kept on control diet C and offspring received the IF-high diet solely during the specific period from PND 21 until PND 50 (CHC group).

In total 14 experimental treatment groups (n=6-12 animals/group) were available to study the impact of IF on estrogen sensitivity and DNA methylation changes in MGs of healthy Wistar rats.

A full list of all experimental animals, interventions and methods used can be found in Table 4 as well as in Table 5.

3.2.4 August Copenhagen Irish rats

A second animal experiment was performed with August Copenhagen Irish (ACI/segHsd) rats at the Technical University in Dresden by the group of Prof. Dr. Günter Vollmer.

This rat strain represents an animal model susceptible to develop MG tumors solely after prolonged exposure to pregnancy levels of exogenous E2 [161]. Recently, Shull et al. developed a novel animal model based on the ACI rat and exhibit the same susceptibility towards breast cancer but promising less co-morbidity associated with pituitary hyperplasia [162]. The use of this rat strain can compromise longer studies focusing on breast tumor development and might be of first choice for future research on mammary cancer prevention. For tumor promotion both ovarian steroids, E2 and progesterone have to be present, since OVX completely abolishes mammary carcinogenesis in ACI rat [163]. Treatment with tamoxifen also reduced the ability of E2 to induce tumors in ACI rats indicating the driving role of E2 in this model of breast carcinogenesis [164, 165]. Distinct quantitative trait loci, *i.e.*, estrogen-induced mammary cancer (Emca) underlie the breast cancer susceptibility of these rats [166]. Four Emca loci are orthologous to human loci that have been associated with higher mammographic breast density and map to genes involved in cell cycle control and DNA Repair (*e.g.*, Cdkn2 and Rad51), as well as in proliferation (Myc), and drug resistance (MRPs) [167]. In addition, arising tumors show expression of both ER α and Pgr which further supports and emphasizes the close connection of this model to human types of breast cancer.

For the lifelong IF intervention protocols, ACI rats were obtained from Harlan and were randomly assigned to control C or H diets prior to mating (Table 3). In the kinetic arm of the experiment, pregnant dams were kept on allocated diets, thus female offspring were exposed to IF during all

developmental windows of fetal and neonatal, as well as pubertal, and adolescent development until the end of the experimental period (Figure 26). In order to better understand the impact of IF on the kinetics of DNA methylation changes during normal MG development, rats were sacrificed before and after puberty at PND 21 and PND 50±1, during adolescence at PND 81±2, and PND 97±2, as well as in adulthood at PND 180±3.

In the carcinogenesis arm of the study, female ACI rats were randomly divided into two different treatment groups (n=10) in order to monitor the impact of IF on MG carcinogenesis (Figure 47). At PND 45, when the number of TEB is the highest, one group of ACI rats was implanted with a 4 mg E2-containing silicon tube to initiate the tumorigenic process. A second tube was implanted at PND 175 to ensure a consistent E2-release during the experiment, since it has been reported that withdrawal of the promoting agent lead to regression and re-differentiation of the pre-neoplastic lesions [168]. Tumor development was monitored by palpation. At the termination of the experiment, tumors were excised, tumor multiplicity was monitored, and tumor size and weights were recorded. Detailed information on animal handling, experimental conditions, and E2-releasing implants is published in Möller et al. 2016 (under revision).

A full list of all experimental animals, interventions and methods used can be found in Table 4 as well as in Table 5.

Table 4: Summary of all experimental animals, interventions and methods used.

Rat strain	Diet	Birth	Dissection	-E2 [n]	+E2 [n]	Start of E2 treatment [PND]	Duration of E2 treatment [d]	Sampling [PND]	Genome wide method	Candidate gene approach	Gene expression	Protein expression
Wistar	CC	03/ 2012	06/ 2012	8	9	94	3	97	-	MA	RT-qPCR ^a	IHC ^b
Wistar	CHC	03/ 2014	06/ 2014	7	7	94	3	97	-	MA	RT-qPCR ^a	
Wistar	CHC	05/ 2013	08/ 2013	6	9	94	3	97	-	MA	RT-qPCR ^a	
Wistar	LL	01/ 2012	04/ 2012	6	7	94	3	97	-	MA	RT-qPCR ^a	IHC ^b
Wistar	MC	10/ 2011	11/ 2011	6	6	94	3	97	MClp-Seq	MA	RT-qPCR ^a	IHC ^b
Wistar	MH	10/ 2011	11/ 2011	12	12	94	3	97	MClp-Seq	MA	RT-qPCR ^a	IHC ^b
Wistar	HH	11/ 2011	02/ 2012	6	7	94	3	97	-	MA	RT-qPCR ^a	IHC ^b
ACI	C, H	07/ 2013	04/ 2014	20	20	45	240	285	RRBS	MA	RT-qPCR ^a	IHC ^c
ACI	C, H	11/ 2011	05/ 2012	12	-	-	-	180	-	MA	RT-qPCR ^a	IHC ^c
ACI	C, H	11/ 2011	03/ 2012	12	-	-	-	97	-	MA	RT-qPCR ^a	IHC ^c
ACI	C, H	11/ 2011	02/ 2012	12	-	-	-	81	-	MA	RT-qPCR ^a	IHC ^c
ACI	C, H	01/ 2012	03/ 2012	12	-	-	-	50	RRBS	MA	RT-qPCR ^a	IHC ^c
ACI	C, H	01/ 2012	02/ 2012	12	-	-	-	21	RRBS	MA	RT-qPCR ^a	IHC ^c

C: control diet, L, M and H: Isoflavone enriched diets (68, 232 and 503 ppm), PND: post natal day, RRBS: Reduced representation bisulfite sequencing, MA: MassARRAY, RT-qPCR: Real time quantitative PCR, IHC: Immunohistochemistry

^a: performed by Karin Klimo, technician in the group of Dr. Clarissa Gerhäuser, DKFZ Heidelberg,

^b: performed by the group of Prof. Dr. Dr. Patrick Diel, DSHS Cologne,

^c: performed by the group of Prof. Dr. Günter Vollmer, TU Dresden.

Table 5: Summary of ACI rats and obtained tissues from E2-induced MG carcinogenesis (PND 285), interventions and methods used.

Tissue	Diet	Birth	Dissection	[n]	Start of E2 treatment [PND]	Duration of E2 treatment [d]	Sampling [PND]	Genome-wide method	Candidate gene approach	Gene/Protein Expression	Epithelial enrichment
MG	C, H	07/13	04/14	20	-	-	285	RRBS	MA	RT-qPCR ^a , IHC ^b	-
MG/T	C, H	07/13	04/14	17	45	240	285	-	MA	RT-qPCR ^a , IHC ^b	-
T	C, H	07/13	04/14	24	45	240	285	RRBS	MA	RT-qPCR ^a , IHC ^b	-
MG	C, H	10/11	06/12	12	-	-	240	-	MA	-	LCM
DCIS	C	10/11	06/12	3	60	180	240	-	MA	-	LCM

MG: mammary gland, MG/T: Mammary gland with no palpable tumor, T: Tumor, DCIS: Ductal carcinoma *in situ*, C: control diet, L, M and H: Isoflavone enriched diets (68, 232 and 503 ppm), PND: post natal day, RRBS: Reduced representation bisulfite sequencing, MA: MassARRAY, RT-qPCR: Real time quantitative PCR, IHC: Immunohistochemistry, LCM: laser capture microdissection

^a: performed by Karin Klimo, technician in the group of Dr. Clarissa Gerhäuser, DKFZ Heidelberg,

^b: performed by the group of Prof. Dr. Günter Vollmer, TU Dresden.

3.2.5 DNA, RNA and protein isolation

Material	
Name	Manufacturer
Agilent RNA 6000 nano kit	Agilent Technologies, Palo Alto, USA
AllPrep RNA/DNA Mini Kit	Qiagen, Hilden, Germany
Isopropanol	Sigma-Aldrich, Taufkirchen, Germany
Chloroform	Sigma-Aldrich, Taufkirchen, Germany
Ethidium bromide	Sigma-Aldrich, Taufkirchen, Germany
Lambda DNA/HindIII Marker	Thermo Fischer Scientific, Waltham, USA
Proteinase K solution	Qiagen, Hilden, Germany
RNase A	Sequenom, San Diego, USA
RNase-free DNase Set	Qiagen, Hilden, Germany
RWT Buffer	Qiagen, Hilden, Germany
Qubit® RNA 6000 Nano Kit	Invitrogen, Life Technologies, Carlsbad, USA
Qubit® dsDNA HS Assay Kit	Invitrogen, Life Technologies, Carlsbad, USA
TRIzol reagent	Sigma-Aldrich, Taufkirchen, Germany
β-mercapto-ethanol	AppliChem, Darmstadt, Germany

Non standard instruments	
Name	Manufacturer
Agilent 2100 BioAnalyzer	Agilent Technologies, Palo Alto, USA
Micro-Dismembrator S	Sartorius, Göttingen, Germany
Nanodrop spectrophotometer (ND1000)	PeqLab, Erlangen, Germany
Qubit® 2.0 Fluorometer	Invitrogen, Life Technologies, Carlsbad, USA

DNA, RNA, and protein was isolated from fresh frozen tissue using the Qiagen AllPrep DNA/RNA kit and TRIzol reagent. The manufacturers' protocol was modified in order to increase DNA/RNA yield and quality as well as fitting the requirements of the usage of TRIzol. All used buffers were obtained from Qiagen.

3.2.5.1 Tissue processing

Fresh frozen tumors and MG were pulverized using the Sartorius Micro-Dismembrator S while shaking flasks were cooled with liquid nitrogen whenever possible. For isolation 20–50 mg of tissue powder were lysed with 1 ml RLT Plus buffer, supplied with 10 µl β-mercapto-ethanol (β-ME). After complete lysis of tissue, samples were centrifuged for 3 min at 10,000 rpm in order to remove non-soluble material. For DNA binding residual liquid was transferred to an AllPrep DNA spin column in

500 μ l portions and flow through was collected. The AllPrep DNA spin column was washed with 500 μ l buffer AW1 in order to remove β -ME, and DNA isolation was carried out as described in the DNA isolation section (see 3.2.5.2). Flow through was further processed by adding of 1 ml TRIzol reagent and vortexed. After incubation for 5 min at room temperature, 200 μ l chloroform were added followed by vortexing and incubation for 15 min at room temperature. Phase separation was achieved by centrifugation for 15 min and 13,000 rpm at 4 °C. The colorless, aqueous phase, containing the RNA, was transferred to a new 2 ml reaction tube, mixed with 500 μ l isopropanol, and stored at -20 °C overnight. RNA isolation was performed as described in the RNA isolation section (see 3.2.5.3). The red, organic phase was used for protein isolation as described in the protein isolation section (see 3.2.5.4).

3.2.5.2 DNA isolation

DNA was bound to AllPrep DNA spin columns as described in the tissue processing section. DNA was eluted in 40 μ l PBS and 10 μ l RNase A was added. RNA was digested for 30 min at 37 °C, and the eluate was transferred to the AllPrep DNA spin column again, following RNase A on-column digestion for additional 15 min at 37 °C. Columns were washed using buffer AW1 and flow through was pipetted on the columns again to ensure complete DNA binding. In order to remove proteins, 20 μ l proteinase K solution was added to 60 μ l buffer AW1 and protein digestion was carried out on-column for 30 min at room temperature. Proteins were washed away with buffer AW1 twice followed by two final washing steps with buffer AW2. In order to remove residual contaminants from buffer AW1 columns were inverted during AW2 washing steps. Columns were dried of residual ethanol by centrifugation for 5 min and 13,000 rpm and DNA was eluted using 50 μ l buffer EB. The eluate was pipetted back on the column in a second elution step in order to concentrate DNA.

Quantification of DNA was done using spectrophotometry at 260 nm and additionally by Qubit® dsDNA BR Assay Kit. DNA quality was checked by loading 200 ng DNA and 3 μ l Lambda DNA/HindIII Marker onto a 0.8 % TBE agarose gel containing 4 μ l/100ml ethidium bromide for visualization (120 V, 90 min).

3.2.5.3 RNA isolation

The RNA containing aqueous phase was stored at -20 °C overnight as described in the tissue processing section. Protein digestion was achieved by adding 80 μ l proteinase K solution and incubation for 30 min at room temperature. In order to ensure RNA and especially miRNA binding to columns, 400 μ l buffer RWT was added and mixed briefly by inverting the tubes. The entire solution was transferred to an AllPrep RNA spin column in 700 μ l portions. After washing the columns with 400 μ l buffer RWT, remaining DNA was digested on-column using 10 μ l DNase I in 70 μ l buffer RDD

(RNase-free DNase Set) for 14 min at room temperature. Columns were washed using 500 µl buffer RWT and flow through was pipetted on the AllPrep RNA spin columns in order to ensure complete RNA binding. Columns were washed twice with buffer RPE. Columns were inverted in both washing steps in order to remove residual contaminants from buffer RWT. Columns were centrifuged dry (5 min, 13000 rpm) and RNA was eluted using 30 µl RNase-free water. The eluate was pipetted back on the column in a second elution step in order to concentrate RNA. All centrifugation steps were carried out at 4 °C.

RNA quality was checked on an Agilent RNA 6000 nano chip and the concentration was measured by spectrometry at 260 nm. Purity was estimated from 260/230 nm and 260/280 nm ratios.

3.2.5.4 Protein isolation

Protein was precipitated from the red, organic phase as described in the tissue processing section by adding 1 ml isopropanol and incubation for 1 h at room temperature. Protein was pelleted by centrifugation for 10 min and 13,000 rpm at 4 °C. Protein pellets were washed twice using 2 ml wash solution (0.3 M guanidine hydrochloride in 95 % ethanol) for 20 min. Between washing steps, washing solution and protein pellets were centrifuged for 5 min and 13,000 rpm at 4 °C. After the final washing step, protein pellets were stored in 100 % ethanol at -20 °C.

3.2.6 Methyl-CpG Immunoprecipitation and Next generation Sequencing

Material	
Name	Manufacturer
Agilent high sensitivity DNA kit	Agilent Technologies, Palo Alto, USA
EDTA	Sigma-Aldrich, Taufkirchen, Germany
MgCl ₂	Sigma-Aldrich, Taufkirchen, Germany
MinElute Purification Kit	Qiagen, Hilden, Germany
NaCl	Sigma-Aldrich, Taufkirchen, Germany
NP-40	Genaxxon Bioscience, Ulm, Germany
QuantiTect SYBR Green PCR kit	Qiagen, Hilden, Germany
Qubit® dsDNA HS Assay Kits	Life Technologies, Carlsbad, USA
Tris-HCl	Sigma-Aldrich, Taufkirchen, Germany

Non standard equipment

Name	Manufacturer
12 x 24mm round-bottom glass tubes	Covaris, Woburn USA
Agilent 2100 BioAnalyzer	Agilent Technologies, Palo Alto, USA
BioFuge Fresco centrifuge	Heraeus, Hanau, Germany
DiaMag protein A-coated magnetic	Diagenode, Liège, Belgium
Heraeus Megafuge 1.0R centrifuge	Thermo Scientific, Rockford, USA
LightCycler480	Roche, Penzberg, Germany
S220 Focused-ultrasonicator	Covaris, Woburn USA
SX-8G IP-Star® Automated System	Diagenode, Liège, Belgium
Qubit® 2.0 Fluorometer	Invitrogen, Life Technologies, Carlsbad, USA

3.2.6.1 High salt and single fraction elution protocol

In order to detect differentially methylated regions (DMRs) in normal mammary tissue of healthy adult female Wistar rats, MCIp experiments were performed as described previously [169, 170] with minor modifications. 2-5 µg high quality genomic DNA was ultrasonicated for 9 min at 4 °C, 10 % duty cycle, intensity 5 and 200 cycles/burst to fragment sizes of roughly 150 bp. Selective binding of methylated DNA to the methyl binding domain of the MBD2-Fc protein bound to magnetic protein A-coated beads enabled the separation of low, moderate and highly methylated DNA. Therefore, 60 µg MBD2-Fc protein and 40 µl of magnetic beads were washed in buffer A (300 mM NaCl, 0.1 % NP-40, 20 mM Tris-HCl pH 8.0, 2 mM MgCl₂, 0.5 mM EDTA) and mixed with 2-5 µg sonicated DNA.

An engineered protocol for the Robot SX-8G IP-Star enabled an automatic washing, bead incubation, and DNA fragment separation according to their methylation status by sequential elution with increasing NaCl concentrations using buffers B-F (buffer B: 400, C: 500, D: 550 and F: 1000 mM NaCl, with remaining components as for buffer A). For the high salt elution protocol (HS) the high affinity fraction eluted with highest salt concentration F was used for subsequent library preparation followed by NGS. This fraction was enriched for highly methylated DNA fragments of CpG-dense regions and CGIs.

In order to enrich for medium and high methylated DNA fragments, a single fraction elution protocol (SF) was used. DNA was washed with low salt buffer B to remove non- and low-methylated DNA fragments from MBD2-Fc protein and the magnetic beads. A second elution with buffer F containing the highest NaCl concentration eluted all DNA fragments from MBD2-Fc protein. Therefore, SF was enriched for high and less CpG-dense regions with a medium and high degree of DNA methylation [81, 85].

Eluates were desalted using the MinElute Purification Kit according to manufacturer's instructions. DNA elution with 10 µl EB twice led to a final volume of 20 µl enriched DNA. For size confirmation, final library quality was checked for unwanted products on an Agilent high sensitivity DNA chip and double-stranded DNA concentration was determined by the Qubit® dsDNA HS Assay.

3.2.6.2 Control for enrichment of methylated DNA by qPCR

In order to check for effective enrichment, 2 µl of spike-in DNA was added to the ultrasonicated DNA. This control DNA was derived from *Arabidopsis thaliana* and was artificially methylated using M.SssI before it was mixed in an equimolar ratio with unmethylated control (1 pg/µl). Elution profiles of the internal controls as well as for the imprinted gene Mest were measured by real time quantitative PCR (RT-qPCR). For primer sequences see 3.1.3.2.

The Lightcycler 480 system and the QuantiTect SYBR Green PCR Kit were used for qPCR. In brief, the following protocol was used: 1.4 µl DNA from the different fractions + 7 µl qPCR Master Mix (3.5 µl QuantiTect SYBR Green 2x, 0.2 µl forward primer (10 µM), 0.2 µl reverse primer (10 µM) and 1.7 µl nuclease-free water) were analyzed with the following program: 15 min initial denaturation at 95 °C, 45 cycles of 15 s at 94 °C, 30 s at 60 °C. Analyses were always run in duplicates. Serial dilutions of ultrasonicated but non fractioned DNA (1:10, 1:50, 1:100 and 1:1000) were used as basis for a standard curve and enrichment profile calculation. Samples were submitted for sequencing (50 bp, single-end reads) on a HiSeq 2000 instrument (Illumina Inc., San Diego, USA). Using the NebNext chemistry (NEB, Ipswich, USA) sequencing libraries were prepared according to the manufacturer's protocols by the DKFZ Genomics and Proteomics Core facility.

3.2.6.3 Bioinformatics data mining

Raw sequencing reads of all animals per diet group were pooled and aligned to the reference genome (rn5, <http://genome.ucsc.edu>) using the BWA aligner [128]. Duplicates and low quality reads (MAQ score of <20) were filtered out using SAMtools [171]. The MEDIPs package of R was used to measure the read coverage of each individual CpG enriched by MCIp. Read counts were normalized to 1 million reads. DMRs (200bp) were generated and annotated using the Homer suite of tools [172]. Promoter regions (2kb upstream and 500 bp downstream of TSS) were downloaded from the UCSC genome browser. Additionally, gene expression data from the gene expression omnibus (GSE38060) was used to further evaluate the candidate list for biological relevance [173]. Putative DMRs were selected for high coverage, high fold change and low q-values. Hyper- and hypomethylated DMRs refer to the comparison of MH intervention groups with the MC control group. Data processing was performed by a set of Perl (<http://www.perl.org/>) and R (<http://www.r-project.org/>) scripts customized by Dr. Lei Gu and Dr. Manuela Zucknick.

3.2.7 Reduced representation bisulfite sequencing - Library preparation

Material	
Name	Manufacturer
5 M NaCl	Sigma-Aldrich, Taufkirchen, Germany
Agilent high sensitivity DNA kit	Agilent Technologies, Palo Alto, USA
dNTP set, 100 mM each	NEB, Ipswich, USA
EB buffer	Qiagen, Hilden, Germany
EZ DNA Methylation-Direct™ Kit	Zymo Research, Irvine, USA
GoTaq qPCR Master Mix	Promega, Mannheim, Germany
Klenow exo-polymerase	NEB, Ipswich, USA
Methylated and unmethylated control oligos	Christoph Bock, CEMM, Vienna, Austria
MspI	NEB, Ipswich, USA
NEB Buffer #2	NEB, Ipswich, USA
PEG 8000	Sigma-Aldrich, Taufkirchen, Germany
Pfu Turbo Cx hotstart DNA polymerase	Agilent Technologies, Palo Alto, USA
Qubit® dsDNA HS Assay Kit	Life Technologies, Carlsbad, USA
Quick ligase	NEB, Ipswich, USA
SPRI AMPure XP magnetic beads	Beckman Coulter, Krefeld, Germany
TruSeq DNA Sample Preparation LT Kit	Illumina Inc., San Diego, USA.

Non standard equipment	
Name	Manufacturer
Agilent 2100 BioAnalyzer	Agilent Technologies, Palo Alto, USA
MagnetoPURE-Micro	Chemicell, Berlin, Germany
Qubit® 2.0 Fluorometer	Invitrogen, Life Technologies, Carlsbad, USA

3.2.7.1 MspI digest, end repair and A-Tailing

In order to generate a reduced representation of the rat genome, MspI, a restriction endonuclease (C↓CGG) which is insensitive to CpG methylation, was used to generate a predictable number of DNA fragments with at least one analyzable CG in each fragment, including highly redundant fragments from microsatellite DNA. Therefore, 100 ng genomic DNA was digested using 20 U MspI and 3 µl 10x NEB Buffer #2 in a total volume of 30 µl for 14 h at 37°C.

The digestion reaction was mixed with 5 U Klenow exo- polymerase, 1 µl dNTP mix (10 mM dATP, 1 mM dCTP and 1 mM dGTP) and 0.1 ng unmethylated and methylated control oligos. Incubation for 20 min at 30 °C filled in the 3'- and 5'-overhangs of the MspI-digestion fragments. A further incubation for 20 min at 37°C was used to add a single A nucleotide to the 3'-terminal ends which

prevents the fragments from ligating to one another and is required for the subsequent adapter ligation. The reaction was inactivated by heating to 75°C for 20 min. Addition of unmethylated and methylated control oligos facilitates subsequent calculations of bisulfite conversion rates.

3.2.7.2 Adapter Ligation

The addition of adapters is a basic principle of sequencing random DNA fragments. These short sequences allow hybridization of the DNA fragments onto a flow cell and enable barcoding of samples to mix multiple DNA libraries together into one sequencing lane. The universal and the indexed adapter have a very short complementary sequence which is composed of twelve nucleotides only, followed by a number of nucleotides which are hanging off [174]. Illumina standard adapters were prepared in 1:50 dilutions in resuspension buffer (provided in the TruSeq DNA Sample Preparation LT Kit) and were forwarded for denaturation at 95 °C for 15 min. A subsequent cool down to 70 °C for 15 min and a step wise cooling protocol to 25 °C over a period of 25 min allowed complementary oligos to anneal and to form “floppy” overhangs. A single T nucleotide on the 3'-terminal end of the adapters serves as a complementary overhang for the A-tailed DNA fragments. 2.5 µl adapters were ligated for 20 min at 25 °C using 1 µl Quick ligase and 37.5 µl Quick Ligase buffer, followed by an inactivation step at 65 °C for 10 min.

3.2.7.3 Cleaning up using SPRI Beads

Adapter ligated fragments were then cleaned by AMPure XP magnetic beads recovering fragments between 150 bp and 1000 bp in length. AMPure XP beads were removed from 4 °C storage at least 30 min prior to usage and were vortexed until they were well dispersed. As total DNA amount was below 100 ng, beads were diluted 1:5 with crowding reagent (25 ml 5 M NaCl, 20 ml 50 % w/v polyethylene glycol 8000, 5 ml nuclease-free water).

0.75 volumes of beads were added to the reaction and mixed gently by pipetting up and down 10 times until beads were resuspended homogeneously. Excessive foaming was prevented by pipetting 2/3 of the volume, only. The mixture was incubated at room temperature to allow binding to the beads and then placed on a magnetic stand for 15 min each. Aspiration of the cleared solution and washing of the beads twice with 200 µl freshly prepared 80 % ethanol removed all residual contaminants, *i.e.*, primers, dNTPs and unbound DNA fragments. Beads were removed from the magnetic stand and were air-dried at room temperature. Over-drying of beads was avoided by adding 25 µl EB for DNA elution as soon as remaining ethanol was fully evaporated. Again, beads were gently resuspended by pipetting up and down 10 times, incubated at room temperature and then placed on a magnetic stand for 15 min each. Eluates, containing the RRBS libraries were transferred into a new tube and stored at 4 °C for further analysis or at -20 °C for long term storage.

3.2.7.4 Quality control qPCR

In order to ensure that end repair and adapter ligation were efficient, a semi-quantitative PCR was performed as quality control. Obtained Ct-values were used for quantifying RRBS libraries to enable homogenous multiplexing of six samples into one sequencing library.

The Lightcycler 480 system and GoTaq qPCR Master Mix were used for qPCR. In brief, the following protocol was used: 1 μ l undiluted library + 19 μ l qPCR Master Mix (10 μ l Master Mix 2x, 2 μ l PCR primer cocktail (PPC, 25 mM each), 7 μ l nuclease-free water) were analyzed with the following program: 3 min initial denaturation at 98 °C, 25 cycles of 15 s at 95 °C, 30 s at 60 °C, 30 s at 72 °C and a subsequent elongation step for 5 min at 72 °C. Analyses were always run in duplicates. Ct-values, calculated with the Roche Lightcycler Software Release 1.5.0 using the Abs quant/2_{nd} derivative max option served both as a quality control (Ct <14) and to calculate the volumes for library pooling.

First, all libraries were sorted by Ct-values in ascending order and six consecutive samples were pooled. The volume of the sample used for pooling with maximum Ct-value (Ct_{max}) was 17 μ l. Difference in Ct-values (Δ Ct) and fielding volumes (v_{sample}) for pooling for the residual five sample were calculated using the following equations (see example below):

$$\Delta Ct = Ct_{\text{max}} - Ct_{\text{sample}}$$

$$v_{\text{sample}} = 17 \mu\text{l} \times 2^{-\Delta Ct}$$

Samples were pooled in combinations of six according to the calculated volumes in a 2 ml Eppendorf tube (Table 6). Libraries with identical adapter numbers in one pool were switched into the next pool, since adapters were used for identification of a particular library. The total volume of the pooled library was adjusted to 33 μ l by adding 2.5 volumes AMPure XP beads and subsequent clean up as described before.

Table 6: Library sorting according to Ct values for multiplexed sequencing.

Library	Ct ₁	Ct ₂	Ct _{average}	Δ Ct	V _{Sample} [μ l]
1	8.94	8.75	8.85	0.71	10.4
2	8.95	8.93	8.94	0.62	11.1
3	9.26	9.15	9.21	0.35	13.3
4	9.44	9.18	9.31	0.25	14.3
5	9.61	9.33	9.47	0.09	16.0
6	9.56	9.55	9.56	0.00	17.0

3.2.7.5 Bisulfite conversion

In order to maintain methylation information, DNA samples were subjected to bisulfite conversion using the DNA Methylation Direct Kit according to manufacturer's recommendation with major modification. Starting the protocol at section II, the pool of 6 libraries was incubated with 117 μ l of CT Conversion Reagent in an Eppendorf Mastercycler for 20 cycles of 1 min at 95 °C and 10 min at 60 °C with a subsequent cooling step for 10 min at 4 °C. Trouble shooting and sequencing of several bisulfite conversion protocols revealed that 16 cycles of 15 sec at 95°C and 60 min at 50°C using the standard Zymo Research methylation Kit increased bisulfite conversion rate and should be used in future analyses (see Figure 34).

For desulphonation, 600 μ l of M-Binding Buffer was added to a Zymo-Spin™ IC Column. Bisulfite-converted library pools were loaded onto a Zymo-Spin™ IC Column and were mixed by inverting the column several times. All Zymo-Spin™ IC Columns were washed with 100 μ l M-Wash Buffer in order to remove denaturing detergences. 200 μ l M-Desulphonation buffer was added to the columns and library pools were incubated for 30 min at room temperature. Columns were washed twice using 200 μ l M-Wash Buffer before bisulfite converted RRBS libraries were finally eluted by adding 20 μ l of M-Elution Buffer to the column's membrane. The eluate was pipetted back on the column in a second elution step in order to concentrate DNA for immediate enrichment PCR.

Bisulfite treatment of DNA leads to a chemical deamination of unmethylated cytosines (C) to uracil (U). In subsequent PCR reactions, all uracils (from unmethylated Cs) are amplified as thymines, whereas only methylated Cs are amplified as cytosines, allowing discrimination of unmethylated and methylated Cs at single CpG resolution. Achieving a consistent C to U conversion of unmethylated DNA is crucial for obtaining reliable quantitative methylation data.

3.2.7.6 Enrichment PCR

The bisulfite-converted DNA was then enriched by PCR amplification using the Pfu Turbo Cx Hotstart DNA polymerase, a proof-reading PCR enzyme that does not stall when it encounters uracil.

To minimize the bias caused by amplification, a semi-quantitative PCR was performed prior to the enrichment PCR to determine the lowest possible number of PCR cycles for enrichment. The Lightcycler 480 system and GoTaq qPCR Master Mix were used for qPCR. In brief, the following protocol was used: 1 μ l undiluted bisulfite converted library pool + 19 μ l qPCR Master Mix (10 μ l Master Mix 2x, 2 μ l PPC (25 mM each), 7 μ l nuclease-free water) were amplified with the following program: 3 min initial denaturation at 98 °C, 25 cycles of 15 s at 95 °C, 30 s at 60 °C, 30 s at 72 °C and a subsequent elongation step for 5 min at 72 °C. Analyses were always run in duplicates. Ct-values, calculated with the Roche Lightcycler Software Release 1.5.0 using the Abs quant/_{2nd} derivative max

option served both as a control (Ct <18) and to calculate the optimal cycle number for enrichment which is typically Ct-value minus 1.

For one enrichment reaction, 20 µl of bisulfite converted RRBS library was incubated with 10 U Pfu Turbo Cx hotstart DNA polymerase, 20 µl 10× Pfu Turbo Cx reaction buffer, 20 µl PPC (25 mM each) and 2 µl dNTP mix (25 mM each dATP, dGTP, dCTP and dTTP) and was adjusted to a total of 200 µl by using nuclease-free water. For enrichment of each library pool the reaction was split into eight aliquots of 25 µl each and was incubated in a thermocycler according to the following program: 2 min initial denaturation at 95 °C, Ct – 1 cycles of 30 s at 95 °C, 30 s at 60 °C, 45 s at 72 °C and a subsequent elongation step for 7 min at 72 °C. Analyses were always run in duplicates.

From now on, all steps were performed in post-PCR environment to control for contamination with amplified RRBS libraries.

PCR-amplified library aliquots were combined into one 2 ml DNA LoBind tube and cleaned by 2 volumes of undiluted AMPure XP magnetic beads as described before. Pfu Turbo Cx reaction buffer is highly viscose and often disturbs subsequent library quality control. Therefore, libraries were washed three times with 1 ml freshly prepared 80 % ethanol to ensure a complete removal of the Pfu Turbo Cx reaction buffer and were then eluted into a 1.5 ml DNA LoBind tube using 20 µl EB.

Between all steps of library preparation, samples were kept on ice.

For size confirmation, final library quality was checked for unwanted products on an Agilent high sensitivity DNA chip and double-stranded DNA concentration was determined by the Qubit® dsDNA HS Assay. Each library was sequenced on one (ACI rats PND 21-50, see paragraph 4.3.3) or two lanes (ACI rats PND 285, see paragraph 4.3.5.2) of an Illumina HighSeq2000 machine (50-base single-end reads) and sequencing results are summarized in Supplemental table 3 and Supplemental table 5.

3.2.7.7 Bioinformatics analysis

Raw sequencing reads were aligned to the rat genome sequence using BSMAP [175]. Data processing was performed by custom Python (<http://python.org/>) and R (<http://www.r-project.org/>) scripts [176]. A validated software tool called RnBeads [177] was used for bioinformatic RRBS data mining. This tool presents quantitative methylation data at single CpG site, CGI and promoter level, allows pair- or group-wise statistical comparison of intervention groups for efficient detection of differentially methylated sites (DMS) or DMRs, generates graphical reports of results, and provides an overview of enriched gene ontologies representing the DMRs [177].

3.2.8 Quantitative DNA Methylation analysis by EpiTYPER technology

Material	
Name	Manufacturer
EZ DNA Methylation™ Kit	Zymo Research, Irvine, USA
“T” Cleavage MassCLEAVE reagent kit	Sequenom, San Diego, USA
dATP, dCTP, dGTP, dTTP	Fermentas, St. Leon-Rot, Germany
DNeasy Kit	Qiagen, Hilden, Germany
HotStar Taq DNA polymerase	Qiagen, Hilden, Germany
HotStarTaq DNA Polymerase kit	Qiagen, Hilden, Germany
M.SssI	Fermentas, St. Leon-Rot, Germany
REPLI-g Mini Kit	Qiagen, Hilden, Germany
RNase A	Sequenom, San Diego, USA
Shrimp Alkaline Phosphatase SAP	Sequenom, San Diego, USA

Non standard equipment	
Name	Manufacturer
Mastercycler 384	Eppendorf, Hamburg, Germany
MALDI-TOF Mass Spectrometry	Sequenom, San Diego, USA
MassARRAY Nanodispenser	Sequenom, San Diego, USA
Sequenom EpiTYPER software 1.2	Sequenom, San Diego, USA

Methylation of interesting candidate regions were quantitatively analyzed using MALDI-TOF Mass spectrometry. The method includes amplification by PCR, SAP treatment, *in vitro* transcription, and base-specific cleavage of nucleic acids, as well as separation of fragments according to their mass to charge ratio (m/z).

The principal of the assay is the conversion of unmethylated C to U by bisulfite conversion and subsequently to adenines during an *in vitro* transcription. Methylated cytosines are unaffected and transcribed into guanines. The exchange of guanine to adenine introduces a mass shift of 16 Da for a single CpG site or 32, 48, 60, 72 Da etc. for several CpG sites in one CpG-unit and results in a unique spectra signal in the MALDI-TOF Mass spectrometry distinguishing methylated and unmethylated Cs. Using this system, methylation differences >5 % are detectable by quantifying median methylation of single CpGs or two or more CpGs together (CpG units) (also see section 1.2.2.4).

3.2.8.1 Bisulfite conversion

Bisulfite conversion or bisulfite treatment of genomic DNA is used to maintain methylation information for downstream applications and is considered as the gold standard for quantitative methylome analyses.

500 ng high quality genomic DNA isolated from the Wistar rats were subjected to sodium bisulfite treatment using the EZ DNA Methylation™ Kit, according to the manufacturers' instructions. In brief, 5 µl M-Dilution Buffer was added to the DNA and the total volume was adjusted to 50 µl with nuclease-free water. After an initial incubation step for 15 min at 37 °C, 100 µl of prepared CT Conversion Reagent was added and samples were incubated using the following program: 16 cycles of 15 s at 95 °C, 60 min at 50 °C and a subsequent cooling step for 10 min at 4 °C (Protocol 2, see Figure 34). For desulphonation, 400 µl of M-Binding Buffer was added to a Zymo-Spin™ IC Column. Bisulfite-converted DNA was loaded onto a Zymo-Spin™ IC Column and was mixed by inverting the column several times. All Zymo-Spin™ IC Columns were washed with 100 µl M-Wash Buffer in order to remove denaturing detergents. 200 µl M-Desulphonation buffer was added to the columns and was incubated for 20 min at room temperature. Columns were washed twice using 200 µl M-Wash Buffer before bisulfite converted DNA was eluted twice in 30 µl M-Elution buffer, 1:2 diluted using nuclease-free water and stored at -20 °C.

Genomic DNA isolated from ACI rats was subjected to sodium bisulfite treatment using the DNA Methylation Direct Kit (Zymo Research) as described for library pools in the RRBS section 3.2.7.5.

In order to check for efficient conversion, a subsequent PCR was performed. Primers were designed to bind to the Sall3 gene, a CG-dense region with many putative C conversions in non-CpG contexts. For one reaction, 1 µl bisulfite converted DNA was incubated with 0.2 U µl HotStar Taq DNA polymerase, 0.5 µl 10x HotStar Taq buffer, 0.1 µl 10 µM Forward-Reverse primer mix and 0.1 µl dNTP mix (2.5 mM each dATP, dGTP, dCTP and dTTP) and was adjusted to a total of 5 µl by using nuclease-free water. Reactions were incubated in a thermocycler according to the following program: 15 min initial activation at 94 °C, 48 cycles of 30 s at 95 °C, 1 min at 60 °C, 1 min at 72 °C and a subsequent elongation step for 5 min at 72 °C. Good quality bisulfite converted DNA served as positive and genomic DNA as well as water as negative controls. For visualization the PCR products were loaded onto 2 % TBE agarose gel (120 V, 50 min). Primers annealed, if all unmethylated Cs were fully converted to Us, resulting in an amplification of the Sall3 region and clearly visible DNA bands at around 170 bp.

3.2.8.2 Primer design and optimization

For quantitative DNA methylation analysis, primers were designed to bind specifically bisulfite converted templates by using the EpiDesigner tool from Sequenom. Basically, primers were designed with an annealing temperature ranging from 56 °C to 64 °C with an optimum at 60 °C. Ideally, primer length was between 20 and 30 bp and poly-thymine stretches of 6 or more were avoided. Finally both strands could serve as template since hemimethylation is a negligible and transient event, which exists either during *de novo* methylation when DNMT3a and 3b establish a new methylation pattern or after DNA replication, shortly before DNMT1 is maintaining the parental methylation signal. Finally, primers should amplify a product with a length between 100 and 500 bp, with an optimal length of 300 bp for high quality input DNA. The RSeqMeth package of R was used to analyze amplicons *in silico* for fragmentation patterns and molecular weights. Primers that generated amplicons with an appropriate number of separately analyzable CpGs were selected for PCR amplification. In order to adapt primers for the MassARRAY protocol, reverse primers were tagged with a T7-promoter and 10mer tags were added to the forward primers to counterbalance the melting temperature.

Primers were tested in a PCR in order to find the optimal annealing temperature to reduce off-target amplification and primer-dimer formation. For one reaction, 1 µl template DNA was incubated with 0.2 U µl HotStar Taq DNA polymerase, 0.5 µl 10x Hot Star Taq buffer, 0.1 µl 10 µM Forward-Reverse primer mix and 0.1 µl dNTP mix (2.5 mM each dATP, dGTP, dCTP and dTTP) and was adjusted to a total of 5 µl by using nuclease-free water. Each primer pair was tested at four different annealing temperatures (56 °C, 58 °C, 60 °C and 62 °C) using good quality bisulfite converted DNA as positive and genomic DNA as well as water as negative templates. Reactions were incubated in a thermocycler according to the following program: 15 min initial activation at 94 °C, 48 cycles of 30 s at 95 °C, 1 min at 56 °C, 58 °C, 60 °C or 62 °C, 1 min at 72 °C and a subsequent elongation step for 5 min at 72 °C. For visualization, the PCR products were loaded onto a 2 % TBE agarose gel (120 V, 50 min). Annealing temperatures showing no off-targets in bisulfite converted or genomic DNA were used for quantitative methylation analysis. For a summary of all MassARRAY primers see section 3.1.3.1.

3.2.8.3 Preparation of methylated and unmethylated standard

In order to test whether the designed primers are able to amplify DNA within a dynamic range, standards reflecting different degrees of methylation were included for each PCR amplification.

For methylated control DNA, 15 µg genomic DNA was incubated with 10 µl M.SssI, 13 µl 50x SAM as well as 65 µl 10x buffer and was adjusted to a total volume of 650 µl using nuclease-free water. After 1 h at 37 °C, the methylation reaction was stopped by placing the tubes at 4 °C.

The bacterial CpG Methyltransferase M.SssI is able to methylate all cytosine-residues in unmethylated and hemimethylated DNA in a CpG-context. However, 5fC and 5caC, intermediates of active demethylation are not recognized by M.SssI and will be deaminated in a subsequent bisulfite conversion. This chemical transmission from cytosine to thymine as well as spontaneous deamination of methylcytosine over evolutionary time is measured as unmethylated cytosines and affects the standard DNA that we will not be able to measure 100 % methylation.

Whole genome amplification (WGA) was used to prepare unmethylated control DNA using the REPLI-g Mini Kit. Amplification was performed for 8 reactions in parallel. Denaturation buffer was prepared by adding 35.5 µl nuclease-free water to 10 µl reconstituted buffer DLB. 30 ng genomic DNA was mixed with 5 µl denaturation buffer and was incubated at room temperature for 3 min. Neutralization buffer was prepared by adding 85 µl nuclease-free water to 15 µl stop solution and 10 µl were added to stop the denaturation process. The WGA master mix was composed of 250 µl REPLI-g Mini Reaction Buffer and 8.3 µl REPLI-g Mini DNA Polymerase. Each reaction was mixed with 30 µl master mix and incubated for 10-16 h at 30 °C. Amplification was stopped by a subsequent heat inactivation for 3 min at 65 °C and tubes were stored at 4 °C for further analysis.

M.SssI-methylated and unmethylated WGA DNA was purified using the DNeasy Blood & Tissue Kit from Qiagen. Briefly, 50 µl DNA was mixed with 150 µl nuclease-free water before 200 µl buffer AL and 200 µl ethanol (96-100 %) were added. DNA was bound to DNeasy Mini spin columns by loading the whole mixture onto a column. In total, three samples (50 µl methylated or unmethylated DNA) were loaded onto one column for purification. Columns were washed by adding 500 µl buffer AW1 and 500 µl buffer AW2. In order to remove residual contaminants from buffer AW1, columns were inverted during AW2 washing steps. Columns were dried of residual ethanol by centrifugation for 5 min and 13,000 rpm and DNA was eluted using 2x 70 µl buffer EB. Quantification of DNA was done using spectrophotometry at 260 nm.

In order to prepare a six-point DNA methylation standard, methylated and unmethylated DNA was mixed to obtain a DNA methylation gradient ranging from 0, 20, 40, 60, 80 and 100 % methylation (Table 7). The accuracy of the gradient is dependent on the precise determination of DNA concentration (25 ng/µl). Standard DNA was stored at 4 °C and was subjected to bisulfite conversion within four weeks.

Table 7: Ratios to obtain a 6-point methylation standard.

methylation	0 %	20 %	40 %	60 %	80 %	100 %
unmethylated DNA [μ l]	40	32	24	16	8	0
methylated DNA [μ l]	0	8	16	24	32	40

3.2.8.4 PCR amplification

The region of interest was subjected to PCR amplification using a specific designed primer set and bisulfite converted DNA. For one reaction, 1 μ l template DNA was incubated with 0.2 U μ l HotStar Taq DNA polymerase, 0.5 μ l 10x Hot Star Taq buffer, 0.1 μ l 10 μ M Forward-Reverse primer mix and 0.1 μ l dNTP mix (2.5 mM each dATP, dGTP, dCTP and dTTP) in a 384-well plate and was adjusted to a total of 5 μ l by using nuclease-free water. Reactions were incubated in a thermocycler according to the following program: 15 min initial activation at 94 °C, 48 cycles of 30 s at 95 °C, 1 min at optimal T_M , 1 min at 72 °C and a subsequent elongation step for 5 min at 72 °C. For visualization, 1 μ l of the PCR-product were loaded onto 2 % TBE agarose gel (120 V, 50 min). Sequences of all EpiTYPER primers used can be found are listed in 3.1.3.1.

3.2.8.5 SAP treatment, *in vitro* transcription and desalting

In order to degrade all remaining unincorporated dNTPs, PCR amplified products were subjected to shrimp alkaline phosphatase (SAP) treatment. 0.3 U SAP (1 U/ μ l) and 1.7 μ l nuclease-water was added to each reaction and was incubated for 20 min at 37 °C followed by an heat inactivation step for 10 min at 85 °C.

In a subsequent *in vitro* transcription, PCR amplified DNA was transcribed into single stranded RNA starting from the T7-promoter tag on the reverse strand. Simultaneously, RNase A cleaves the phosphodiester bond at the 3' end of pyrimidine ribonucleotides enabling a U- or C-specific cleavage of the just transcribed RNA fragment. Since ribo-cytosines were replaced by desoxyribo-cytosines in the *in vitro* transcription reaction RNase A cleavage induced a U-specific cleavage pattern of the RNA fragments, only.

For *in vitro* transcription and base-specific cleavage the MassCLEAVE T7 Kit from Sequenom was used. For one reaction, 2 μ l SAP treated DNA was incubated with 22 U μ l T7 RNA&DNA Polymerase, 0.89 μ l 5x T7 Polymerase buffer, 0.22 μ l T Cleavage Mix (contains dCTP, rUTP, rGTP, and rATP), 0.22 μ l DTT (100 mM) and 0.06 μ g RNase A (10 mg/ml) and was adjusted to a total of 7 μ l by using nuclease-free water. Reactions were incubated for 3 h at 37 °C.

For desalting, 20 µl nuclease-free water and 6 mg resin were added to each sample and the 384-well plate was incubated under rotation for 45 min at room temperature. Using the Sequenom Nanodispenser robot, 15 nl of the cleavage products were dispensed on a silicon chip preloaded with matrix and were subjected to MALDI-TOF mass spectrometry analysis.

3.2.8.6 Analysis of quantitative methylation data

Methylation spectra from 1,500-7,000 Da were analyzed using the EpiTYPER software in the recalibrated spectra view. Uncertainty threshold was set to 0.1 and signals were excluded from the analysis, if i: CpG units had a low intensity with less than 100 between signal and noise, ii: CpG units showed silent peaks in the methylated or unmethylated peak, iii: CpG units had salt peaks, overlapping or duplicate peaks, iv: CpG units were uninformative in more than 50 % of the samples, and if v: CpG units showed unexpected values of the corresponding methylation standard. Median amplicon methylation was used to calculate median methylation differences between control and intervention groups.

3.2.9 Quantification of nucleic acids using real time quantitative PCR (RT-qPCR)

Material	
Name	Manufacturer
dNTPs	Fermentas, St. Leon-Rot, Germany
Primers	Sigma-Aldrich, Taufkirchen, Germany
random hexamer primers	Life Technologies, Paisley, UK
Roche Universal Probe Library	Roche Diagnostics, Mannheim, Germany
Roche Universal Probes master mix	Roche Diagnostics, Mannheim, Germany
Superscript III Reverse Transcriptase	Invitrogen, Life Technologies, Paislay, UK

Non standard equipment	
Name	Manufacturer
Heraeus Megafuge 1.0R centrifuge	Thermo Scientific, Rockford, USA
Lightcycler 480 system	Roche Diagnostics, Mannheim, Germany
Roche Lightcycler Software Release 1.5.0	Roche Diagnostics, Mannheim, Germany

RT-qPCR was performed by Karin Klimo.

3.2.9.1 cDNA synthesis

1.5 µg isolated RNA was subjected to reverse transcription (RT) with 1 µl 10 mM dNTPs and 250 ng random hexamers using 200 units of Superscript III Reverse Transcriptase per reaction according to

the cycling conditions in the manufacturer's protocol. After heat inactivation of the transcriptase, cDNA was directly diluted 1:10 with water for qPCR analysis.

3.2.9.2 Real time quantitative PCR

Probe-compatible RT-qPCR primers were designed with the Roche Universal Probe Library Assay Design tool. Whenever possible, intron-spanning primers were preferred to exclude amplification of residual genomic DNA in the samples for mRNA quantification. UPL primer sequences are summarized in 3.1.3.3.

The Lightcycler 480 system and Probes master mix were used for qPCR. In brief, the following protocol was used for a 384-well plate: 2.5 μ l diluted cDNA, 3.5 μ l Probes Master, 1 μ l 10 μ M Forward-Reverse primer mix, 0.05 μ l universal probe with the following program: 10 min initial denaturation at 95 °C, 45 cycles of 10 s at 95 °C, 20 s at 55 °C, 1 s at 72 °C. Ct-values were calculated with the Roche Lightcycler Software Release 1.5.0 using the Abs quant/ 2_{nd} derivative max option. All primers were checked for PCR-efficiencies above 1.94, calculated relative to a cDNA standard ranging from 1:10, 1:50, 1:100, and 1:1000 dilutions.

For cDNA transcript quantification, expression of target nucleic acids was calculated relative to the housekeeping genes Hypoxanthine-guanine phosphoribosyltransferase (Hprt), TATA box binding protein (TBP) and β -actin using the $\Delta\Delta$ CT-method [178].

Statistical analysis was performed as described in section 3.2.12.

3.2.10 Laser capture microdissection (LCM)

Material	
Name	Manufacturer
500 μ l AdhesiveCap, opaque adhesive filling	Zeiss, München, Germany
MembraneSlide 1.0 PEN	Zeiss, München, Germany
QIAamp DNA Micro Kit	Roche Diagnostics, Mannheim, Germany

Non standard equipment	
Name	Manufacturer
ZEISS P.A.L.M. MicroBeam	(Zeiss, München, Germany)
P.A.L.M. Robo V4 software	(Zeiss, München, Germany)
Qubit® dsDNA HS Assay Kit	Agilent Technologies, Palo Alto, USA

MGs were dissected from ACI rats at PND 180 at the TU Dresden by the group of Prof. Dr. Günter Vollmer (see section 3.2.1). Cryosections were collected on PEN-membrane covered slides and stained with hematoxylin and eosin (H&E) in a standard operating procedure at TU Dresden by the group of Prof. Dr. Michael Muders.

In order to enrich for mammary epithelial cells from the cell mix present in a bulk MG, non-contact LCM was performed using the ZEISS P.A.L.M. MicroBeam system. This two-step procedure consists of cold ablation and catapulting of the area of choice into adhesive cap tubes without any contact to the specimen and against gravity. Focus and laser energy were optimized for each process using a pulsed UV laser (355 nm, 2 ns, max. 100 Hz) with an energy output at the outlet opening (objective) <math><20 \mu\text{J}</math> (2 mW). Mammary epithelial cells were focused using an upright widefield microscope (Axio Imager.Z1) and a ZEISS FLUAR 5x/0,25 Air objective lens. The check point function of the Robo V4 software was used to control for correct catapulting and the cutting area was calculated in terms of mm^2 . For healthy MG between 3-5 mm^2 and roughly 1 mm^2 from DCIS samples were used for DNA isolation.

Genomic DNA from healthy animals was isolated using the QIAamp DNA Micro Kit according to the manufacturer's recommendations with minor modifications. Briefly, 15 μl Buffer ATL and 10 μl Proteinase K were mixed and added to the collected laser microdissected sample in the adhesive cap. For protein digestion, samples were incubated upside down for >12 h at 56 °C before tubes were centrifuged and 25 μl of additional Buffer ATL was added. For DNA binding, 50 μl of Buffer AL as well as 50 μl ethanol (96-100 %) were added, mixed and the liquid was transferred to a QIAamp MinElute column. Columns were washed by adding 500 μl buffer AW1 and 500 μl buffer AW2. In order to remove residual contaminants from buffer AW1 columns were inverted during AW2 washing steps. Columns were dried of residual ethanol by centrifugation for 3 min and 13,000 rpm and DNA was eluted using 20 μl buffer EB. The eluate was pipetted back on the column in a second elution step in order to concentrate DNA. Quantification of DNA was done using Qubit® dsDNA HS Assay Kit. Bisulfite conversion for MGs was performed as described in section 3.2.7.5.

Laser captured specimens from tumors were directly subjected to bisulfite conversion using the DNA Methylation Direct Kit (Zymo Research) according to the manufacturer's instructions with minor modification. Briefly, samples were incubated with Proteinase K for >12 h at 56 °C and bisulfite conversion was directly performed using the cell lysate as described in 3.2.7.5.

3.2.11 Immunohistochemistry (IHC)

Protein expression of proliferating cell nuclear antigen (PCNA) and progesterone receptor (Pgr) MGs of Wistar rats was performed using IHC by the group of Prof. Dr. Dr. Patrick Diel at the German Sports University Cologne as previously described [179] (see section 4.3.1.3).

Protein expression of marker of proliferation Ki67 and Pgr in MGs of ACI rats was performed by the group of Prof. Dr. Günter Vollmer at the TU Dresden as previously described [180] (see sections 4.3.1.3 and 4.3.5.1).

3.2.12 Statistical analysis

3.2.12.1 Wistar rats and developmental effects in ACI rats

Methylation values were analyzed as median amplicon methylation. Spearman correlation coefficient rho was calculated to assess pairwise correlation of methylation levels between genes. Beta-regression was used to compare methylation levels between diet and ER groups. Likelihood-ratio test was used to test for global differences between multiple groups. Contrast tests based on the beta-regression model were used to make pairwise group comparisons. Interaction term was tested to identify E2-dependent group effects. Spearman correlation coefficient rho was calculated to assess correlation between methylation and gene expression levels. Corresponding 95 % confidence intervals based on Fisher's z-transformation were calculated. For group comparisons, gene expression values were log₂-transformed if beneficial. Linear regression was used to compare expression levels between groups. F-test was used to test for global differences between multiple groups. Contrast tests based on the linear regression model were used to make pairwise group comparisons. Interaction term was tested to identify E2-dependent group effects.

3.2.12.2 E2-induced breast carcinogenesis in ACI rats

A linear mixed model was used to assess differences in DNA methylation and mRNA expression levels between diets and tissue types. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. A model with diet and tissue effect was fitted. In a second model interaction term between diet and tissue was included and tested. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction in order to control the false discovery rate. Calculations were performed with R package nlme.

4. Results

The aim of the present project was to investigate the impact of IF on DNA methylation and estrogen responsiveness *in vivo*. A short-term rat uterotrophic assay was conducted to characterize the E2-modulating ability of different IF exposure scenarios. Furthermore, a long-term E2-induced rat mammary carcinogenesis experiment was performed to investigate the mammary cancer preventive potential of IF in a daily dietary intervention protocol.

4.1 Effects of IF exposure on hormone responsivity in healthy Wistar rats

Previous studies have demonstrated the potential effects of IF on several endocrinological parameters and a variety of cellular signaling pathways. In order to examine the impact of IF-intervention on DNA methylation, we investigated MG tissue of rats that were maintained for distinct time periods on different diets containing IF at increasing doses (Figure 8).

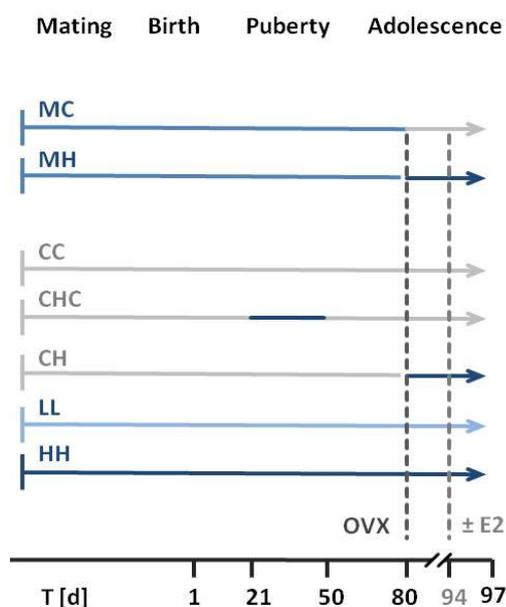


Figure 8: Experimental outline of the rat uterotrophic assay in Wistar rats.

Periods of exposure to IF-free control diet C (3 ppm IF) are indicated by grey lines, exposure to IF-enriched diets (low (L), medium (M), and high (H) IF diets containing 68, 232 and 503 ppm IF aglycone equivalent, respectively) are indicated by different shades of blue. All animals were OVX at PND 80, stimulated with either E2 or vehicle from PND 94 for three days and sacrificed at PND 97. Group names indicate dietary IF exposure before and after OVX.

After OVX at PND 80 and decline of endogenous hormones for two weeks, a three-day rat uterotrophic assay was conducted to analyze whether the IF intervention modulated E2 sensitivity of the adult MG. Effects regarding proliferation, ER-dependent and -independent gene expression as well as DNA methylation were investigated at the day of sacrifice. Additionally, onset of puberty and uterine wet weights (UWW) were monitored as a measure of hormone responsivity. Main results for groups CC, LL and HH have recently been published, including further parameters [160].

4.1.1 Onset of puberty and rat uterotrophic assay (Results provided by Blei/Diel, DSHS Cologne)

For the rat uterotrophic assay the soy extract was added to an IF-depleted control diet (C, 3 ppm IF) leading to average IF aglycone equivalents of 68 ppm for IF-low diet (L) and 503 ppm for IF-high diet (H). Overall intake of the C, L and H diets resulted in a average daily exposure to IF aglycone equivalents of 0.2 mg/kg bw, 4.9 mg/kg bw and 36.6 mg/kg bw, respectively (for details see [160]).

Lifelong IF exposure at low or high concentrations (LL, HH), but also when started at PND 21 (CHC) significantly influenced the onset of puberty as specified by an earlier vaginal opening by about 1.5 days ($p=0.0073$, one way ANOVA) (Figure 9). IF intervention without E2 stimulation had no effect on UWW, independent of the time point of exposure and dose. However, treatment with E2 led to a significant increase of UWW in all diet groups ($p<0.0001$). Interestingly, animals exposed to high IF levels only during the period of hormonal decline (after OVX, group CH) were more sensitive to E2 than animals on control diet (CC) and UWW increased by 12-fold, whereas exposure to H diet only during puberty (group CHC) slightly reduced responsivity and increased UWW by only 5-fold, suggesting a preventive effect of early IF exposure (not significant). Notably, lifelong low, medium and high levels of dietary IF slightly increased E2-induced UWW compared to the control group. This effect was not observed when subsequently to OVX animals were exposed to H diet (group MH), which lead to a less sensitive response to E2 stimulation and resulting in slightly lower UWW (not significant).

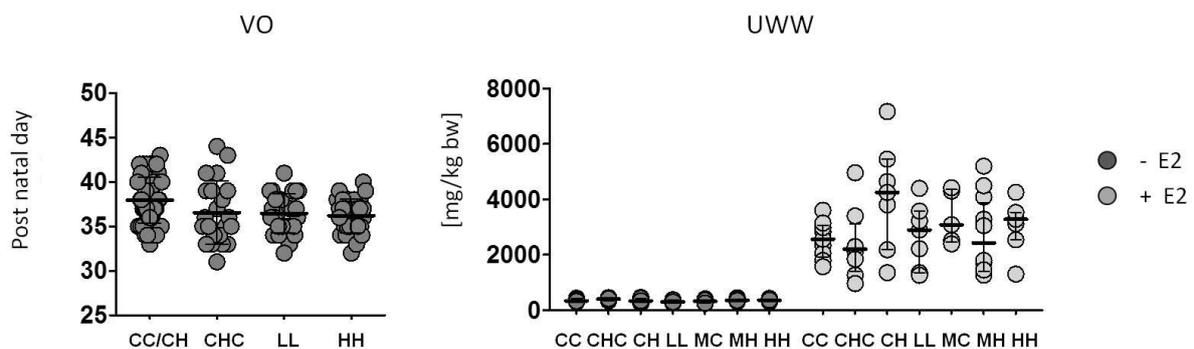


Figure 9: Effects of IF intervention on vaginal opening and uterus responsivity to E2 stimulation.

Each dot in the scatter plots represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97) for vaginal opening (VO) and uterus responsivity to E2 stimulation measured by uterus wet weight (UWW). Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

4.1.2 Estrogenic and proliferative response measured by reference genes (Results provided by Karin Klimo)

In order to investigate the effects of the three day E2 challenge on the MG, we determined the estrogenic response by measuring changes in mRNA expression of selected reference genes for E2 responsiveness (*i.e.*, Pgr, Greb1 and C3), for proliferation (*i.e.*, Ki67 and PCNA), for TFs involved in E2 signaling (*i.e.*, Gata3 and FoxA1), as well as hormone receptors (*i.e.*, AR, Esr1, Esr2 and Vdr) and Krt8 as a marker for epithelial cells (Figure 10 and for statistical analysis Table 8). Gene expression levels were measured in all dietary groups except for MC and MH, for which no RNA was available.

As expected, expression of Pgr, as a prototypical estrogen responsive gene, was low in normal MGs and got massively upregulated upon post OVX E2 challenge (Figure 10A). This E2-induced effect was not as pronounced for C3 and Greb1. Here, mRNA expression was significantly modified by the different IF treatments particularly due to the CH and HH diet. For C3 the post OVX IF intervention (CH group) induced C3 expression to levels observed after E2 treatment, whereas, lifelong high IF levels (HH) showed a rather anti-estrogenic response and prevented C3 expression. Significant interactive effects of lifelong IF levels and E2 exposure further increased gene expression for Greb1 in a dose-dependent manner and suggested an estrogenic mode of action.

Similarly, the proliferation marker PCNA illustrated an E2 response leading to increased mRNA expression (borderline significant $p=0.0549$) although IF intervention alone as well as in combination with E2 treatment affected mRNA expression of Ki67 and PCNA to a greater extent (Figure 10B and Table 8). Interestingly, short term exposure to high IF levels strongly modified estrogen response in the proliferation markers in an opposing manner. For PCNA, estrogenic response was increased by pubertal (CHC) and post OVX (CH) high IF dose whereas mRNA expression levels for Ki67 were reduced.

Although effects were not as pronounced, expression of hormone receptors Vdr and Esr1 was clearly affected by E2 treatment (for Esr1 borderline significant $p=0.0577$) (Figure 10C). Esr1 was massively upregulated after a short term post OVX IF exposure (CH) and Vdr was additionally affected by a pubertal IF exposure (CHC), increasing mRNA expression of the non E2-treated group to levels observed after E2 treatment. Notably, E2 responsive gene expression was enhanced after both pubertal and short term post OVX high IF exposure.

GATA3, a TF involved in E2 signaling and an important regulator for luminal differentiation did not show a pronounced E2 response but interesting expression patterns due to different IF exposures (Figure 10D).

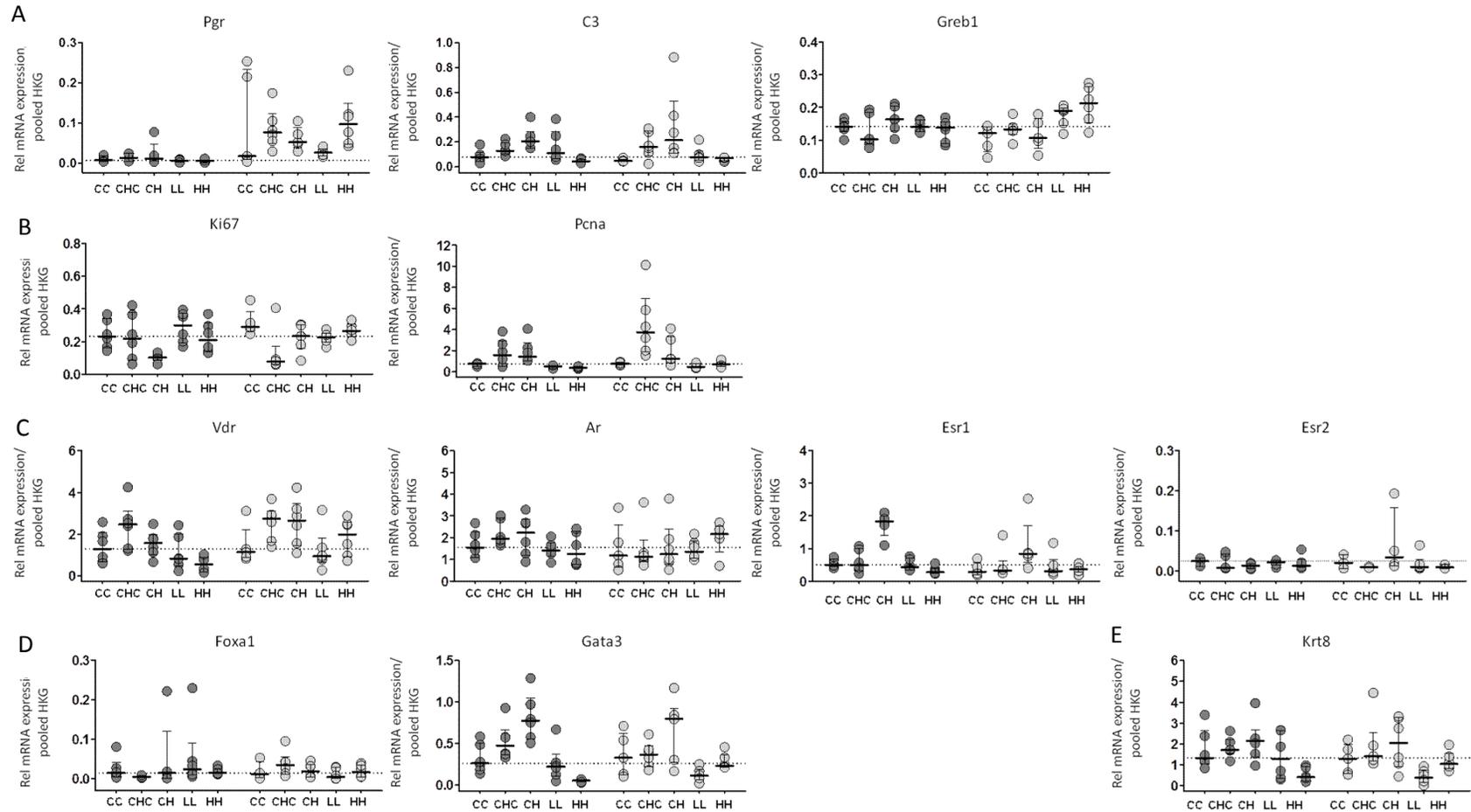


Figure 10: Estrogenic and proliferative response detected by reference gene expression in Wistar rats

For markers of A: E2 responsiveness, B: proliferation, C: hormone receptors and D: for TFs involved in E2 signaling, as well as E: Krt8, as a marker for epithelial cells. Each dot in the aligned dot plots represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Median expression per group is indicated by a horizontal line, with vertical lines indicating interquartile range. The median expression level of the control group is indicated by a dashed line. For statistical analysis see Table 8. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

While a lifelong IF intervention (group HH) reduced GATA3 mRNA expression, short term high level of IF increased GATA3 expression independent of timing of exposure to OVX (groups CHC and CH).

Krt8 as a marker for epithelial cells was not significantly modified by the E2 challenge (Figure 10E). However, the expression pattern resembled the IF-induced estrogenic pro-proliferative effects by the short term high IF exposure as seen for, *e.g.*, PCNA, Gata3 and Esr1, as well as the preventive E2-repressive action as observed for C3, Gata3 and Vdr, induced by a lifelong IF intervention which was attained in a dose dependent manner (groups LL and HH). Ar, Esr2 and Foxa1 were not affected by any treatment.

Table 8: Statistical analysis of the estrogenic and proliferative response of Wistar rats.

Source of Variation	E2 treatment [p value]	IF intervention [p value]	Interaction [p value]
Ar	0.4546	0.7364	0.4373
C3	0.7713	0.0002	0.4949
Esr1	0.0577	< 0.0001	0.2205
Esr2	0.4669	0.3701	0.1136
Foxa1	0.414	0.8242	0.2386
Gata3	0.5696	< 0.001	0.1365
Greb1	0.4108	0.0351	0.0024
Ki67	0.5689	0.0101	0.0222
Krt8	0.4517	0.0020	0.2251
PCNA	0.0549	< 0.0001	0.0430
Pgr	< 0.0001	0.3133	0.2817
Vdr	0.0275	0.0018	0.2815

Statistical significance of differences was calculated for all reference genes using two way ANOVA with the factors: E2 treatment, IF intervention and interaction between E2 treatment and IF intervention across the dietary groups available (CC, CHC, CH, LL, HH).

4.1.3 IHC in mammary gland tissue (Results provided by Blei/Diel, DSHS Cologne)

In order to evaluate the influence of IF on the estrogenic and proliferative response of MG epithelia cells, protein expression of Pgr (PR) as a sensitive marker for E2 responsiveness and of PCNA a well-established proliferation marker was determined by IHC (Figure 11). This technique enables a quantitative assessment of PR and PCNA protein expression in mammary epithelial cells independent of the bulk of cell types by which they are surrounded.

As previously observed for PR gene expression (see section 4.1.2), no signal for positively labeled nuclei for PR was detected in all groups without E2 stimulation. However, treatment with E2 significantly increased the percentage of positively labeled nuclei to up to 27 % in the LL diet group. A

highly statistical significant effect of the combination of diet and E2 could be observed. Interestingly, low IF levels in diet (group LL) increased the percentage of positively labeled nuclei after E2 treatment by an additional 6 % when compared to the control group CC (21 %), whereas lifelong exposure to high IF levels in diet (group HH) prevented the increase in PR expression and resulted in 10 % positively labeled nuclei.

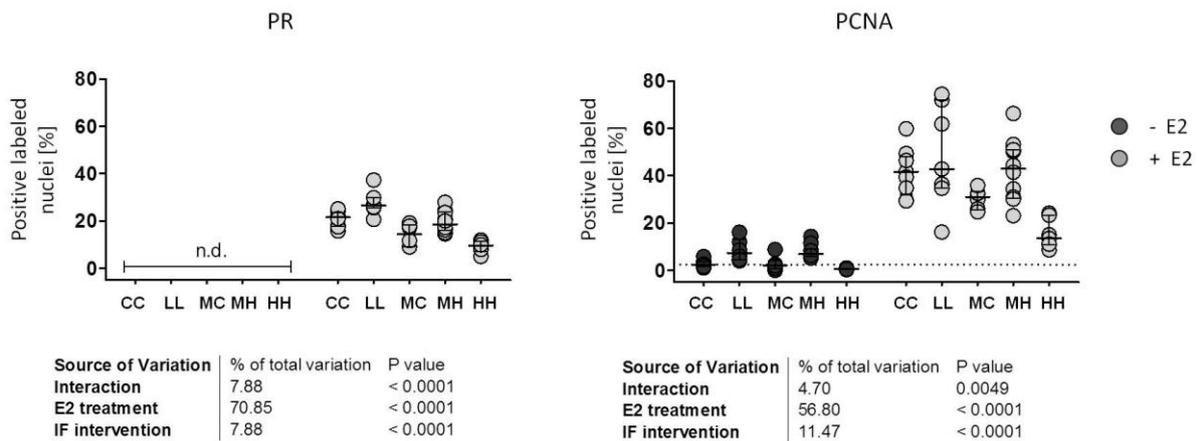


Figure 11: PR and PCNA protein expression levels in response to IF and E2 exposure in Wistar rats.

Protein expression of the E2-responsive PR and the proliferation marker PCNA in the MG determined by IHC. Each dot in the scatter plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. The median expression level of the control group is indicated by a dashed line. Statistical significance of differences was calculated using Two-way ANOVA with the factors: E2 treatment, IF intervention and interaction between E2 treatment and IF intervention. n.d. not detected. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

Interestingly, medium levels of dietary IF pre OVX (group MC) reduced the E2-induced PR protein expression by about one third compared to the control group. This effect was not observed when subsequently to OVX animals were exposed to H diet (MH), which lead to a more sensitive response to E2 stimulation and upregulation of PR protein expression to levels of the control group. Thus, high IF exposure after OVX abolished the preventive effect of a medium pre OVX IF dose, but not of a lifelong high IF dose.

Treatment with E2 and IF alone as well as the combination of both significantly increased the proliferation index in the MG from 0 % in the CC group up to 50 % positively labeled nuclei for PCNA in the LL group. Interestingly, low IF levels in the diet (group LL) elevated PCNA expression also in the absence of E2 (~10 % positively labeled nuclei) whereas high IF levels (group HH) prevented the increase in PCNA expression in response to E2 treatment by two thirds compared to control diet to only 15 % nuclei positively labeled for PCNA. Notably, for PCNA we could observe a similar effect in the MC and MH groups as for PR protein expression. Medium levels of IF prior OVX (MC) reduced

PCNA protein expression to only 30 % positively labeled nuclei. This was reversed when animals were exposed post OVX to H diet (MH), which lead to an increase in PCNA protein expression comparable to levels of the control group. For further details see Blei et al. [160].

Results for both proteins are in line with the relative mRNA expression of these two genes obtained by RT-qPCR (Figure 10).

Overall, these data indicated that dependent on the timing and/or on the dose, exposure to IF quite differentially affected hormone responsivity of the adult MG in Wistar rats.

4.2 DNA methylome mining in normal mammary glands of Wistar rats

4.2.1 Enrichment of high and medium methylated DNA for MClp-Seq

In order to investigate IF-induced changes in DNA methylation on a genome-wide scale, animals of the MC and MH groups were analyzed. These two groups were selected for the analyses because the tissue samples were the first to be available. Methyl-CpG Immunoprecipitation (MClp) was used to enrich for medium and highly methylated DNA fragments, followed by next generation NGS (Figure 12).

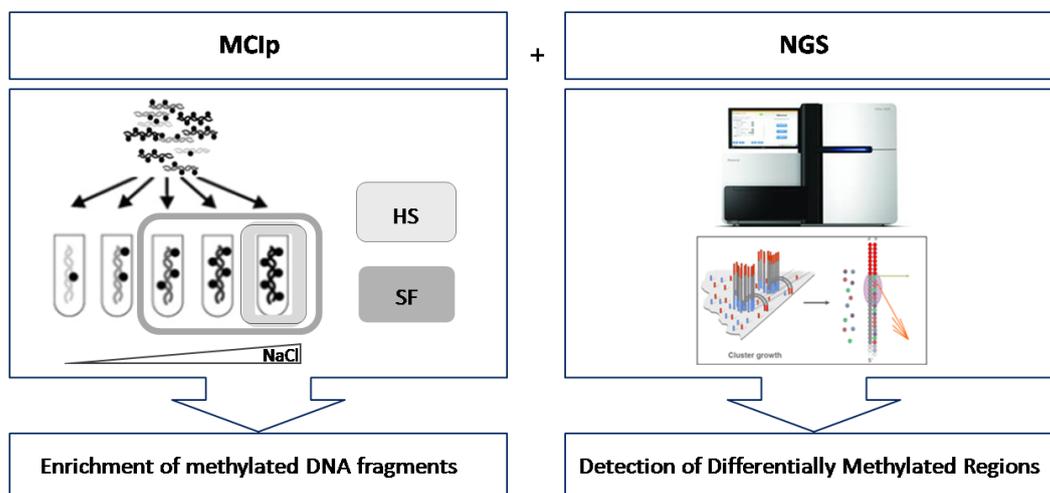


Figure 12: Flowchart of MClp-Seq.

HS, High salt elution; SF, single fraction elution; NaCl, sodium chloride; (pictures partly provided by <http://www.illumina.com>).

Briefly, for enrichment of methylated fragments DNA from groups MC and MH was ultrasonicated and fractionated according to CpG density and methylation status by sequential elution with increasing sodium chloride concentrations. It has been suggested that highly methylated CGIs might be more resistant to demethylation than regions with intermediate levels of methylation [181].

Therefore, we performed two MClp experiments, enriching for either highly methylated DNA fragments only, using the highest salt containing fraction (HS) or for medium and high methylated fragments using a single fraction elution protocol (SF) [75].

4.2.2 Identification of IF-induced differential methylated regions

Raw sequencing reads for samples of all animals per diet group were pooled and aligned to the reference rat genome (release rn5, March 2012). Duplicates and low quality reads were filtered out. Read coverage of each individual CpG and the overall genomic coverage were measured.

4.2.3 Genomic coverage

In total, 17.5 to 18.3 M reads were mapped from the HS elution protocol which covered 24-28 % of all available CpGs in the rat genome at least once (Figure 13). For the SF elution, additionally containing medium methylated fragments, we obtained 45.4 to 71.7 M reads per group which covered 76-85 % of all CpGs.

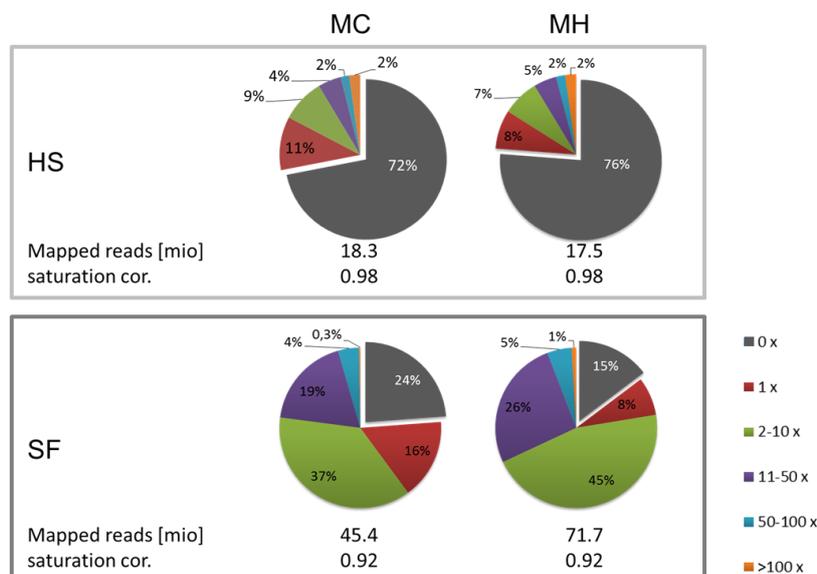


Figure 13: Genomic coverage of MClp-Seq.

Coverage depth per single CpG for A: high salt (HS) and B: single fraction (SF) elution. Data processing was the performed using the MEDIPs working package of R.

Since reads were distributed over more CpGs, the coverage depth per single CpG was lower than in the HS evaluation (saturation correlation 0.98 vs. 0.92) and the number of reads per CpG were lower. Read counts were normalized to 10 M reads. IF-induced DMRs were generated and annotated using the Homer suite of tools [172]. Data processing was performed by a set of custom Perl and R scripts.

4.2.3.1 Relative genomic distribution

Bioinformatics comparison of genome-wide methylation profiles of the MC and MH groups revealed that exposure to distinct IF-containing diets after OVX led to relatively few methylation changes (Figure 14A). Overall, we observed more pronounced hypo- than hypermethylation in HS (3,014 vs. 2,077 DMRs) and slightly more hypermethylation for the SF (296 vs. 1,062 DMRs). The relative genomic distribution of the DMRs was very similar in the HS and SF experiment and for hypo- and hypermethylation, with less than 1 % mapping to regulatory regions (promoters, CpG islands or 5' UTRs), 3-4 % located in exons, and about 90 % accounting for intronic regions, repetitive sequences (LINE, SINE, LTRs, simple repeats and satellites) and intergenic regions.

When comparing hypo- and hypermethylated DMRs identified for HS and SF elution, we detected distinct sets of DMRs with just little overlap between different elution protocols as shown in the Venn diagrams (Figure 14B). This might be partially explained by the lower coverage in the SF experiment and the small number of DMRs.

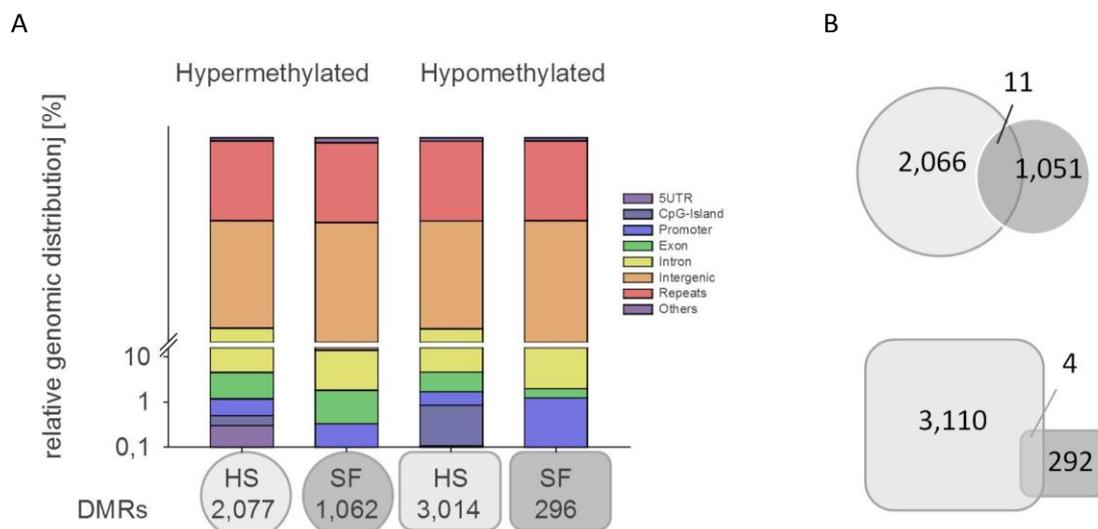


Figure 14: Genomic distribution of DMRs obtained by MChp-Seq.

A: Relative genomic distribution of hypo- and hypermethylated DMRs identified between the MC and MH groups. B: Venn diagrams for overlapping DMRs between HS and SF experiment. (light grey, HS, high salt elution; dark grey, SF, single fraction elution, circles, hypermethylated DMR, squares, hypomethylated DMRs)

4.2.3.2 Overlap and TF motif enrichment

In order to find TF binding sites that might be enriched in the DMRs identified between the MC and MH groups, a *de novo* TF binding motif discovery as well as an enrichment analysis of known TF binding motifs was performed using the Homer suite of tools [172] (Figure 15).

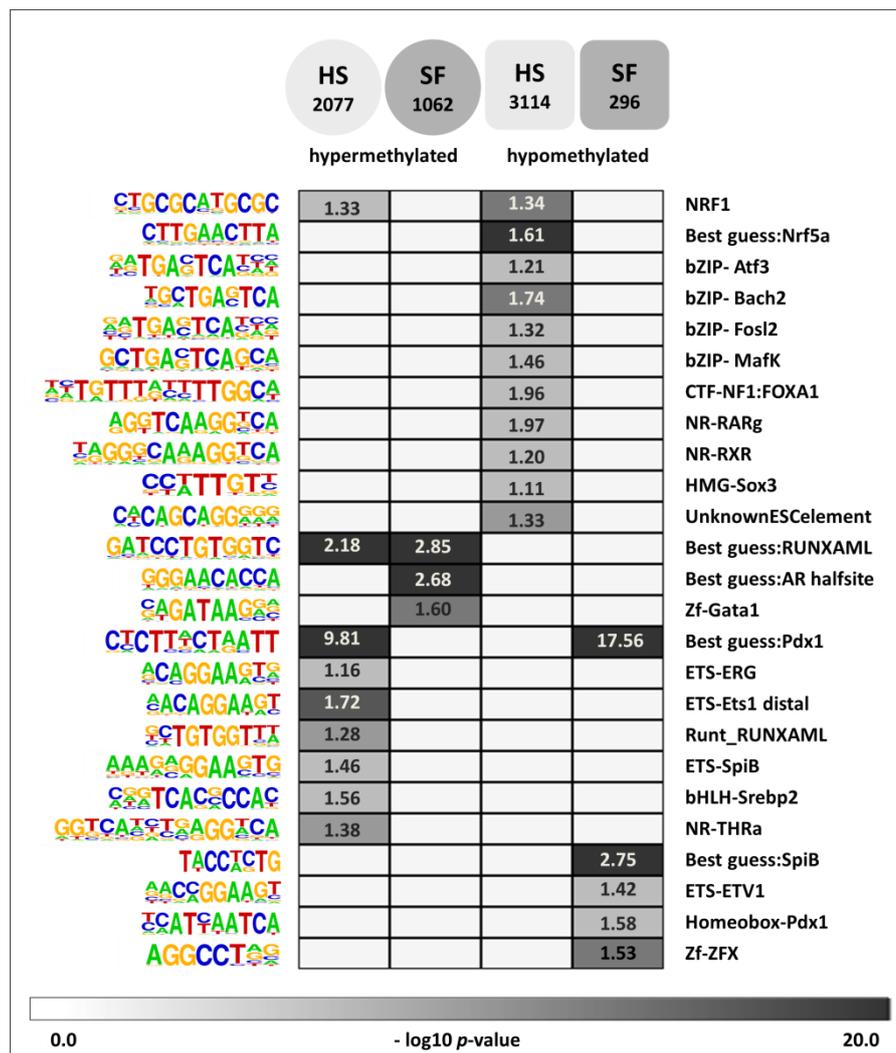


Figure 15: Enrichment of TF binding sites in IF-induced DMRs obtained by MChp-Seq.

Heatmap showing motif enrichment in hypo- and hypermethylated DMRs detected by high salt (HS) and single fraction elution (SF) of MC vs. MH group. Different shades of grey indicate the negative decadal logarithm of the enrichment P-value ranging from 0 (light grey) to 20 (dark grey). Enrichment scores are depicted as numbers in squares and are calculated by dividing the percentage of targets in a given sequence with the percentage of the targets found in background sequences. Only significant motifs in at least one of the four DMR sets are shown.

Distinct results were obtained when analyzing motif enrichment for both datasets, as only little overlap between enriched motifs was detectable between HS and SF. Enrichment was predominantly found for basic leucine zipper (bZIP)-domain TFs (*i.e.*, Atf3, Fosl2 and MafK) in hypomethylated DMRs and motifs of E26 transformation-specific (ETS)-domain TF family members (*i.e.*, ERG, SpiB and Ets1) in hypermethylated DMRs. Both families have been implicated to play important roles in tissue development and cancer progression.

4.2.4 Confirmation of IF-induced DMRs obtained from MClp-Seq using MassARRAY

For validation of DMRs between the MC and MH group observed in the genome-wide methylation analysis by MClp-Seq, quantitative analysis of DNA methylation was performed using mass spectrometry-based EpiTYPER MassARRAY technology. Using this system, methylation differences $\geq 5\%$ are detectable by quantifying median methylation of CpG units.

We selected 14 DMRs with 4-fold enrichment difference between the MC and MH groups, mainly located in gene promoter regions or close to the TSS from both MClp experiments to quantify methylation in all experimental groups (Supplemental table 2).

DMR-associated genes are involved in energy homeostasis and metabolism (Niacr1, Sirt4), DNA repair (Gadd45b), cell cycle regulation (Cdh13), cell signaling and cell proliferation (Wap), cell-cell communication (Cldn4), cell structure (Gsn), estrogen (Esr1 and Cyp2c6v1) and one-carbon metabolism (Aldh1L1) and function as TSGs (PTEN and Wif1) or are frequently down regulated in mammary tumors (Arpp21). Extl1 is highly inducible by E2 in rats [173] but its role in human MG development has not yet been investigated.

For most of the candidate regions we were not able to confirm IF-induced methylation changes between groups MC and MH detected by MClp-Seq which is an enrichment-based methylome profiling technique. MassARRAY analysis revealed median methylation differences below 10% (for detailed group wise comparisons see Supplemental figure 2) with high interindividual variation in methylation levels within dietary groups. Solely for Gadd45b and Cyp2C6v1 we were able to identify a statistical significant effect of the post OVX H IF diet on DNA methylation (Figure 16).

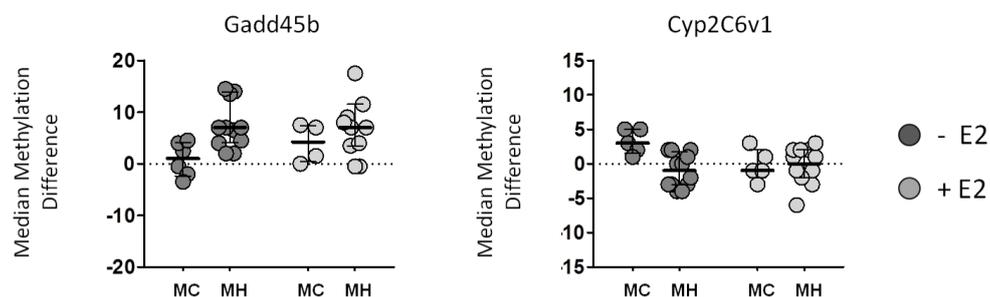


Figure 16: Median methylation differences between MC and MH groups.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. Median methylation of the CC group is indicated as horizontal dashed line and summarized Supplemental table 2. For group wise comparisons of median DNA methylation for all 14 analyzed candidate genes across all dietary groups see Supplemental figure 2. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), medium (M), and high (H) IF diets).

A summarizing figure for median DNA methylation differences in the MC and MH groups relative to the control CC for the other candidate genes can be found in the Supplemental figure 1.

4.2.5 Response to dietary IF is dependent on dose and timing of exposure

In order to investigate rats that were maintained for distinct time periods and diets containing IF at increasing doses, we further assessed DNA methylation in IF-induced DMRs identified by MCIp-Seq, in MG tissues of the remaining experimental groups using MassARRAY technology (Figure 8).

4.2.5.1 Critical time windows

Generally, lifelong exposure rather than a short term treatment regimen is reported to be related to beneficial effects to soy IF. Therefore, we were interested whether the observed effects in DNA methylation were mainly introduced by the pre OVX IF exposure or whether the short term post OVX treatment was the changing trigger.

In order to assess this question, methylation data across several groups were compared dependent on their dietary regime before and after OVX. In principal, groups receiving the same IF dose before OVX were combined and the effects of different IF doses prior to OVX were assessed (Figure 17A). The same approach was used for post OVX comparison. Here, the dietary groups that were exposed to high IF levels after OVX were considered to have the same treatment regime and were compared to groups not receiving any IF post OVX (Figure 17B).

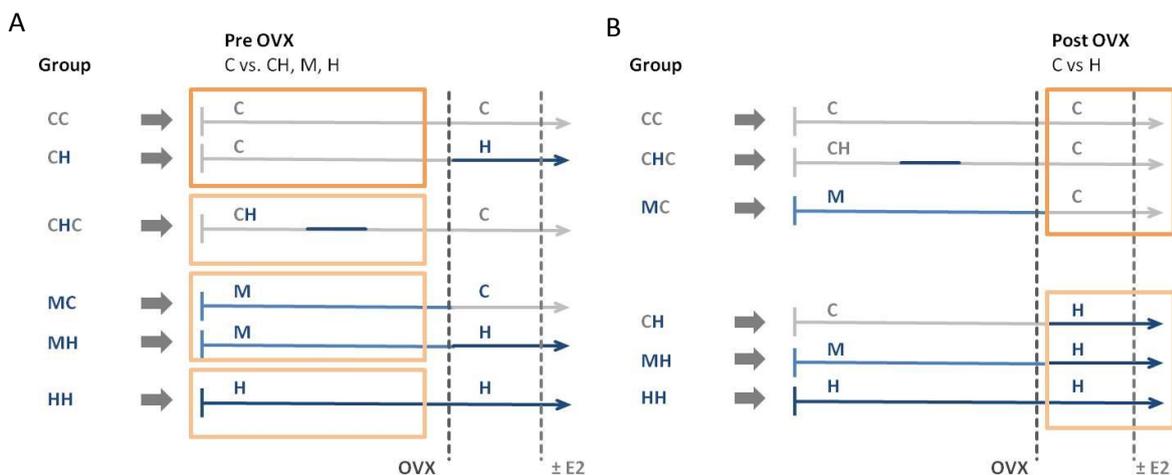


Figure 17: Statistical outline to assess the timing effects of a pre vs. post OVX IF intervention.

Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets). Periods of exposure to C diet are indicated by grey lines, exposure to IF-enriched diets is indicated by different shades of blue. Comparison groups are combined according to their A: pre or B: post OVX dietary regime and are highlighted in rectangles (dark orange: combined control group, light orange: combined group for comparison).

In general, three different comparison groups were defined for the pre OVX effects, receiving either M or H diet before OVX or swapped to H during pubertal phase only (CH). For post OVX comparison a single comparison group was available since either C or H diet was administered after OVX.

Following our statistical analysis more than 75 % of the investigated regions (11 out of 14), DNA methylation changes were already present prior to OVX independent whether a medium or high IF dose was administered, with three regions (Gadd45, Gsn and Sirt4) affected by either M or H diet (Figure 18A).

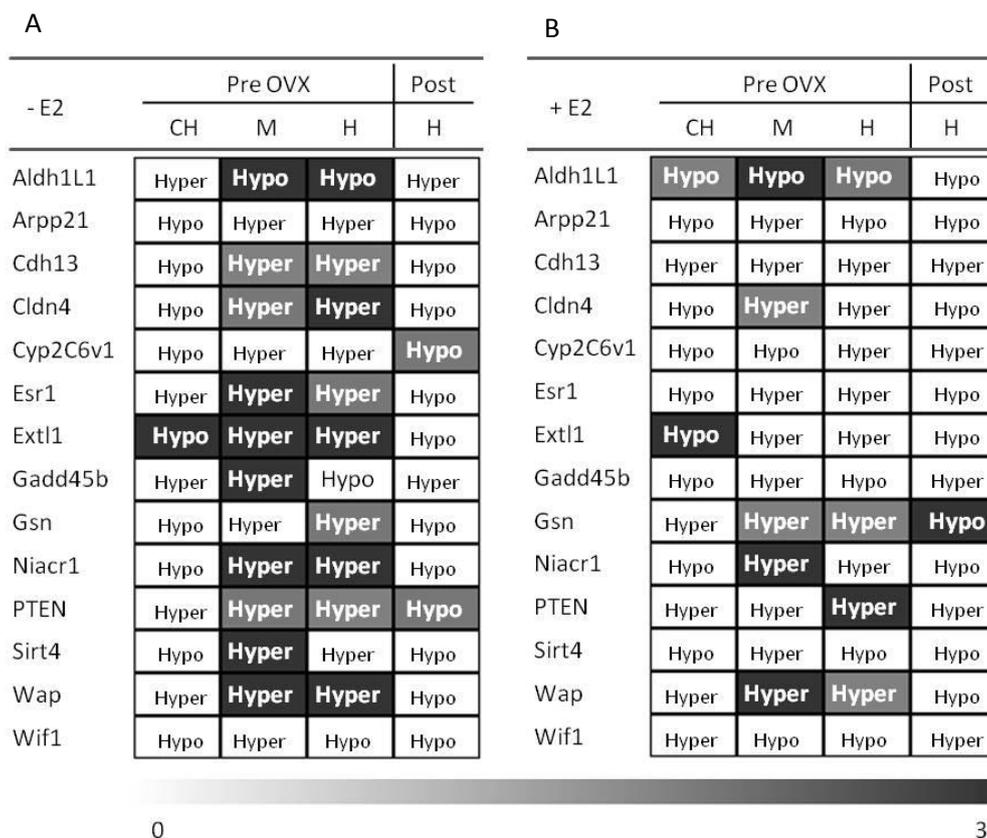


Figure 18: Effects of a pre vs. post OVX IF intervention on DNA methylation.

Statistical analysis across several experimental groups according to their pre and post OVX dietary regime (for details see Figure 17). Likelihood-ratio test was used to test for global differences in DNA methylation between multiple groups. Different shades of grey indicate the negative decadal logarithm of the three distinct significance levels (from light to dark: $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, white = not significant). Text in boxes corresponds to methylation change compared to the combined control group (Hyper: increase and Hypo: decrease of methylation; A: without and B: with subsequent E2 treatment after hormonal decline (PND 94).

Interestingly, in more than 90 % of all comparisons early and lifelong exposure to IF increased DNA methylation when compared with the combined control group. Solely the promoter region of PTEN and Cyp2C6v1 showed a significant effect of the additional post OVX treatment using a high IF dose. Here, a clear hypomethylation of the candidate region could be detected.

In general, effects of pre and post OVX IF treatment were less pronounced when rats were challenged with E2 subsequently after OVX. Although tendencies were less significant, methylation was rather increased due to an early and lifelong exposure to IF and predominantly decreased when IF was applied only or for an additional period during hormonal decline after OVX (Figure 18B).

Exposure to high IF only during the pubertal phase (group CHC) does not seem to be responsible for the effects on DNA methylation seen due to lifelong and early IF exposure. Only *Extl1* and *Aldh1L1* showed a significant change in promoter methylation. However, no clear picture of the direction of DNA methylation changes could be drawn by pubertal exposure to IF. Methylation was either slightly increased or decreased when compared to the combined control group (not significant) and often did not follow any tendency of methylation changes induced by lifelong exposure (Figure 18).

4.2.5.2 Dose aspects of IF intervention

In order to address the question whether the observed timing effects were additionally affected by the different doses of IF, we had a look at dose-effect relationships. First, we looked at all groups exposed to high IF levels post OVX and compared the trends of the different IF doses administered pre OVX. Three different IF diets namely CH, MH and HH were available for comparison with the control group CC. Interestingly, exposure to high IF levels only after OVX (CH group) lowered methylation levels of many of the analyzed regions in comparison to CC and all other dietary groups (Figure 19). With increasing dietary IF to medium and high levels pre OVX, methylation often increased relative to the control group CC, resulting in significant methylation differences between the CH, MH and the HH groups, *e.g.*, for *Niacr1*, *Cldn4*, *Extl1*, *Wap*, *Gsn*, and *PTEN*, as well as *Esr1*, *Gadd45*, and *Sirt4*, with less pronounced effects after E2 treatment (for a summary of statistical analyses see Supplemental figure 2). For *Aldh1L1*, we observed an anti-correlated pattern. Here, we observed strongly decreased DNA methylation in the promoter region in the HH group and increased methylation in the CH group.

When comparing the ability of different pre OVX IF doses to introduce alterations in DNA methylation, we were able to detect significant differences between the CH, MH and HH diet. In 10 out of 14 regions pre OVX methylation changes were induced dose-dependently with a linear trend towards hypermethylation in 8 out of the 10 affected genes (Figure 20A left panel). Only *Gadd45* and *Sirt4* showing a bell-shape-like methylation pattern with the medium IF levels increasing methylation and higher levels decreasing methylation to values comparable with the control group (Figure 19B). We observed a major impact of a medium and high pre OVX IF exposure on DNA methylation with alterations up to 18 % (*Wap*). No significant effects of different pre OVX IF-containing diets were

detected for Arpp21, Cyp2C6v1, Wif1 and Cdh13, although the last was identified as being affected by a pre OVX medium and high IF treatment, previously (Figure 18).

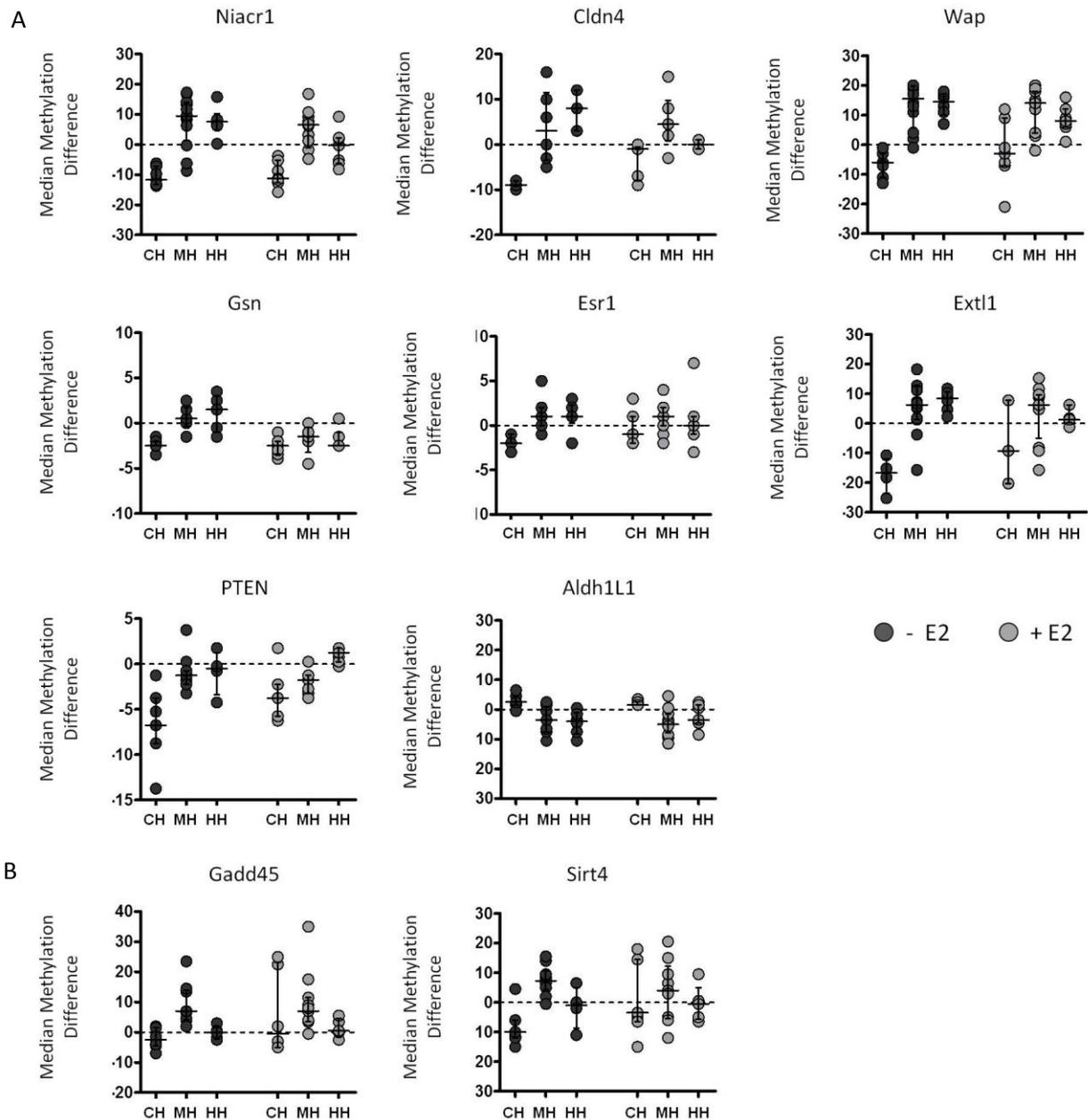


Figure 19: Trends of different pre OVX IF doses (CH, M and H) on median methylation differences.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97), summarizing A: linear and B: U or bell-shape like alterations in DNA methylation. Median values per group are indicated by a horizontal dashed line, vertical lines indicate interquartile range. Median methylation of the CC group is indicated as horizontal dashed lines and summarized in Supplemental table 2. For group wise comparison and statistical calculations of dose-trend-relationships see Supplemental figure 2 and Figure 20, respectively. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), medium (M), and high (H) IF diets).

Next, we were interested whether a lifelong low dose IF treatment (group LL) would be sufficient to induce the effects obtained with medium and high IF diets. We compared the groups LL and HH, two

treatment regimens with a constant lifelong IF exposure, to group CC with no IF exposure. As already described above, lifelong HH diet increased methylation of several genes compared to CC, *i.e.*, *Cldn4*, *Extl1*, *Wap* and *Niacr1*. Notably, low IF levels exhibited opposite effects. Promoter regions were hypomethylated in MGs of rats exposed to LL diet when compared to CC and HH groups and led to a significantly altered DNA methylation profile in a U-shape like dose-dependent manner (Figure 20A right panel and Figure 21A). Exclusively, *Extl1*, *Niacr1* and *Wap* gained methylation at the corresponding promoter regions due to the low IF exposure and a more linear trend towards hypermethylation could be detected (Figure 21B). Neither lifelong low nor high IF levels modified *Arpp21*, *Cdh13*, *Esr1*, *Gadd45b*, and *Sirt4* methylation levels significantly.

Overall, there seems to be a critical IF concentration leading to distinct biological effects since the lower IF concentrations rather oppose the effects observed with high dietary IF levels.

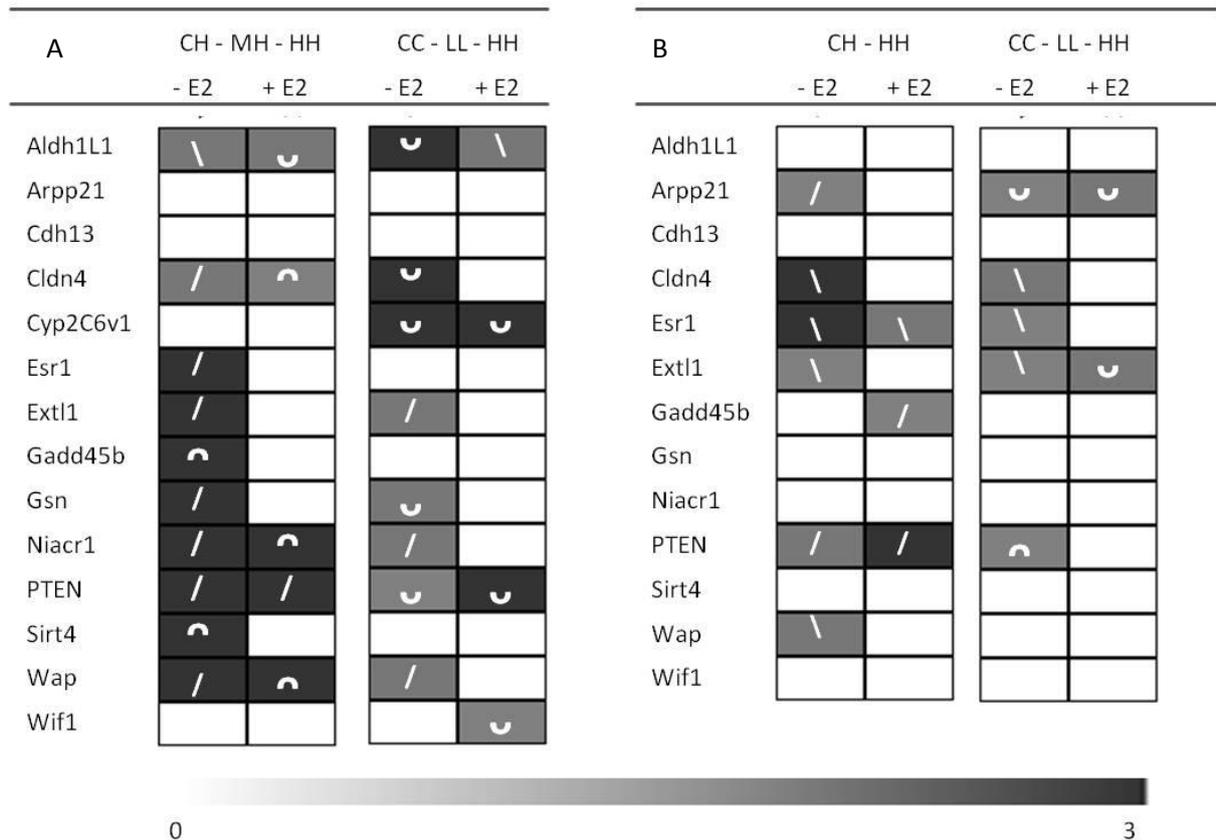


Figure 20: Summary of significant alterations of DNA methylation and gene expression by different IF doses.

Statistical analysis across several experimental groups according to their dosage regime for A: DNA methylation and B: gene expression changes. Different shades of grey indicate the negative decadal logarithm of the three distinct significance levels (from light to dark: $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, white = not significant). Shapes in boxes correspond to changes in median methylation (A) and relative mRNA expression (B) compared to the control group (/: increase, \: decrease, U or bell-shaped) with and without subsequent E2 treatment after hormonal decline (PND 94). n.d.: not determined. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

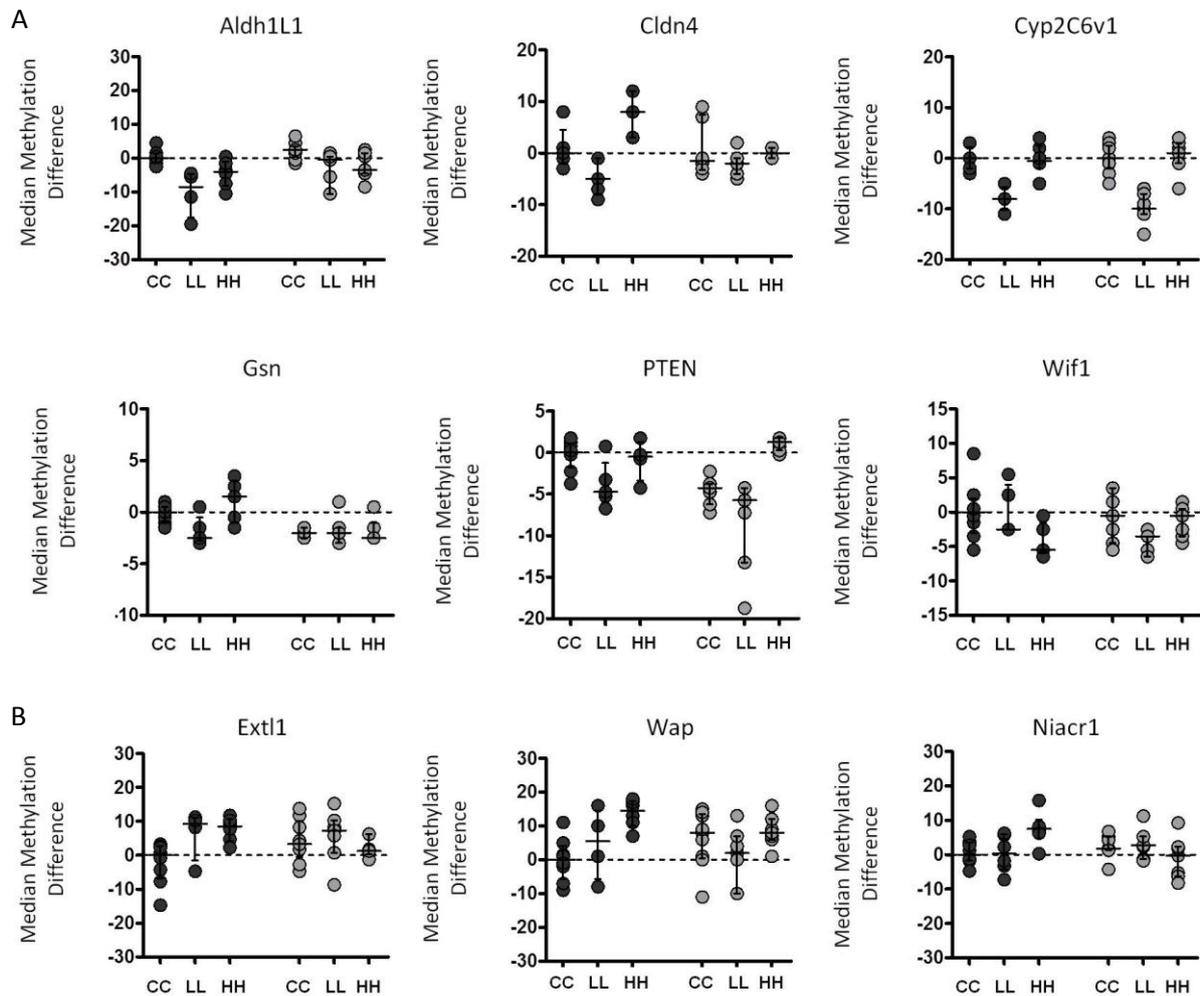


Figure 21: Dose dependent effects of a lifelong IF exposure on median methylation differences.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97), summarizing A: U- or bell-shape like and B: linear alterations in DNA methylation. Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. Median methylation of CC groups are indicated as horizontal dashed lines and are summarized in Supplemental table 2. For group wise comparison and statistical calculations of dose-trend-relationships see Supplemental figure 2 and Figure 20, respectively. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), and high (H) IF diets).

As expected, E2 treatment had little impact on DNA methylation of the investigated regions since we selected particularly IF affected DMRs for quantitative analysis. However, in certain regions a lifelong pre OVX exposure to high IF levels altered the susceptibility to E2 treatment which is depicted as a decrease (*i.e.*, *Cldn4*, *Extl1* and *Wap*) or increase (*Wif1*) in median promoter methylation (Figure 21, Supplemental figure 4). For *Gsn*, promoter hypomethylation was achieved rather by a post OVX H diet (MH, HH). On the other site, a prevention of the estrogenic response towards loss of methylation (*i.e.*, *PTEN*, *Gsn*) or increase in methylation (*i.e.*, *Niacr1*, *Extl1*) due to the lifelong pre OVX M or H diet was detected. Principally, there was only a minor impact of the IF exposure on the

estrogen responsiveness and in more than 70 % this effect was exclusively achieved in the HH group. A summarizing figure for median methylation differences across all experimental groups can be found in the Supplements (Supplemental figure 5).

4.2.6 Effects of an IF intervention are reflected in gene expression patterns (analyses performed by Karin Klimo)

In order to determine whether changes in DNA methylation might be correlated with changes in gene expression, RT-qPCR was performed. Gene expression levels were measured in all dietary groups except for MC and MH, for which no RNA was available. Spearman correlation coefficient rho was computed for the comparison of promoter methylation and gene expression.

Methylation levels of 6 out of 14 genes significantly correlated with expression levels, with correlation coefficient rho in the range of -0.34 to -0.52 for negative correlation of Ext1, Cldn4, Wap, and Esr1 (higher methylation leads to reduced expression) and 0.29 to 0.33 for Gsn and Wif1 with weak positive correlation (Figure 22).

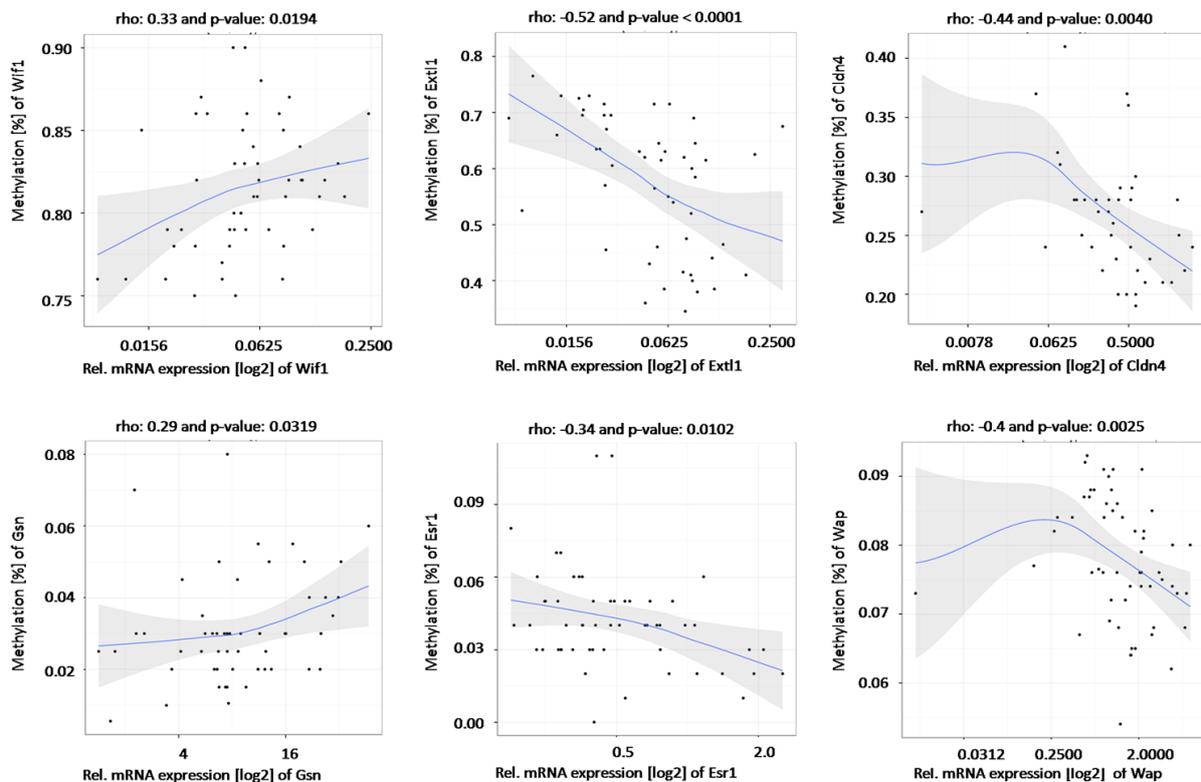


Figure 22: Correlations between DNA methylation and gene expression levels in candidate genes obtained by MCIP-Seq. Each dot represents one sample. Spearman correlation coefficient rho was computed for the comparison of promoter methylation and gene expression. Corresponding 95 % confidence intervals based on Fisher's z-transformation were calculated (grey area). rho: correlation coefficient.

Interestingly, all DMRs for which methylation levels correlated with expression levels were annotated to promoter regions, emphasizing the relationship between promoter methylation and regulation of gene transcription.

Analysis of gene expression confirmed our observation that timing of IF exposure relative to OVX was in many cases a critical determinant for biological effects. Corresponding to the methylation levels, gene expression was oppositely affected by pre vs. post OVX H diet. Hypermethylation due to pre OVX exposure to high levels of IF mainly lead to transcriptional downregulation whereas hypomethylation, predominantly induced by post OVX H diet, increase gene expression (although not consistent for all candidate genes) (Figure 23).

- E2	Pre OVX		Post	+ E2	Pre OVX		Post
	CH	H	H		CH	H	H
Aldh1l1	up	up	down	Aldh1l1	down	up	down
Arpp21	up	up	down	Arpp21	down	up	down
Cdh13	up	down	up	Cdh13	down	down	down
Cldn4	up	down	up	Cldn4	up	down	up
Esr1	down	down	up	Esr1	up	down	up
Extl1	up	down	down	Extl1	up	up	up
Gadd45b	down	up	down	Gadd45b	up	up	down
Gsn	down	up	down	Gsn	down	up	down
Niacr1	down	down	up	Niacr1	up	up	down
PTEN	down	up	down	PTEN	down	up	down
Sirt4	up	down	down	Sirt4	down	down	down
Wap	up	down	up	Wap	up	down	down
Wif1	up	down	up	Wif1	up	down	up



 0 3

Figure 23: Effects of a pre vs. post OVX IF intervention on gene expression.

Statistical analysis across several experimental groups according to their pre and post OVX dietary regime. Different shades of grey indicate the negative decadal logarithm of the three distinct significance levels (from light to dark: $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, white = not significant). Text in boxes corresponds to methylation changes compared to the combined control group (up: increase and down: decrease in gene expression); with (right) and without (left) subsequent E2 treatment after hormonal decline (PND 94)). Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C) and high (H) IF diets).

Exposure to the high IF dose exclusively during the pubertal period exhibited only minor effects on mRNA levels. Consistent with Extl1 hypomethylation, gene expression was significantly up regulated, whereas PTEN expression was reduced in the group exposed to E2 after PND 94 (without a change in

DNA methylation). These results indicate that pubertal exposure to IF might persistently determine transcriptional regulation of these loci.

Gene expression was significantly affected by a short term post OVX vs. lifelong IF exposure detectable between the CH and HH groups (Figure 24). As already indicated in Figure 19, in seven out of 14 genes exposure to high IF levels induced alterations in gene expression dependent on the exposure time. IF exposure only after OVX downregulated Arpp21, PTEN, and Gadd45, whereas Cldn4, Esr1, and Wap mRNA expression levels were induced. For Esr1, PTEN, and Gadd45 IF-induced effects were independent of the E2 treatment, since IF-induced gene expression changes were persistent also after a three day E2 challenge.

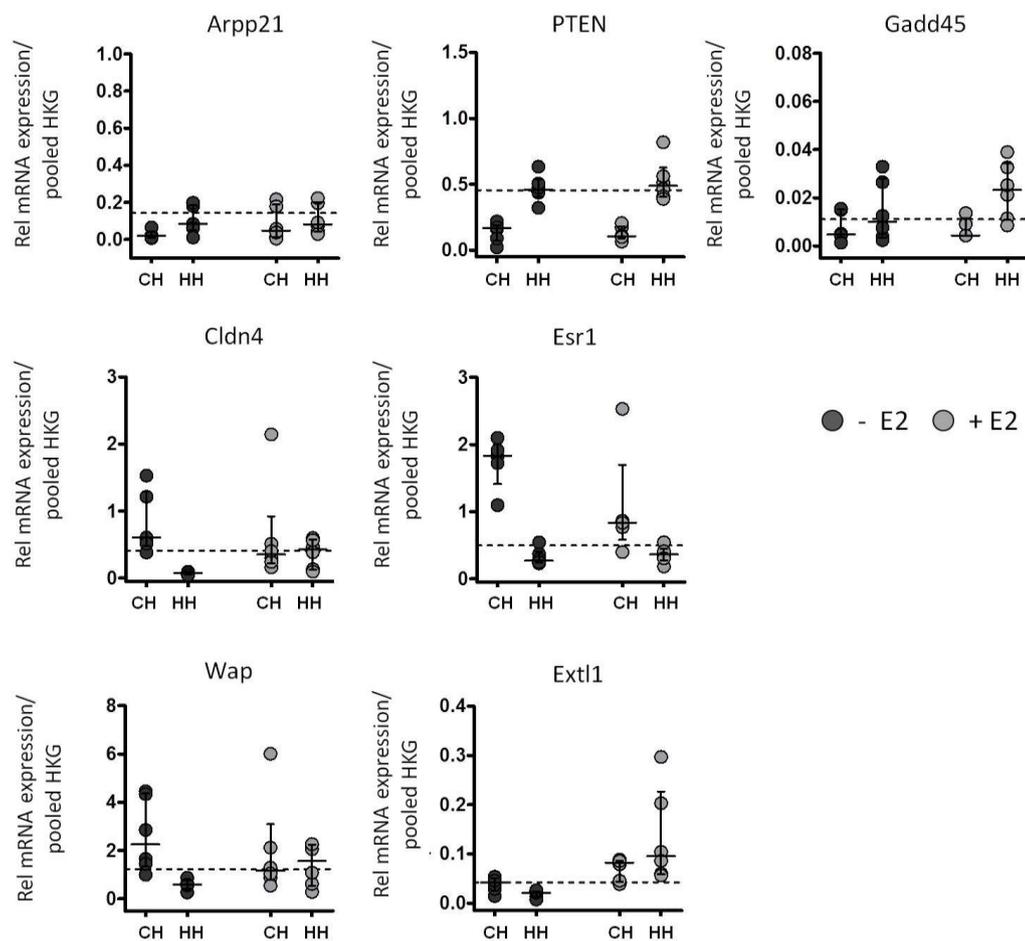


Figure 24: Trends of short term vs. lifelong high (CH and HH) IF doses on gene expression.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Median rel. mRNA expression to pooled housekeeping genes (HKG: Hprt, TBP and β -actin) is indicated as horizontal lines, with vertical lines depicting interquartile range. Median rel. expression of the CC group is indicated as horizontal dashed line. For group wise comparison and statistical calculations of dose-trend-relationships see Supplemental figure 3 and Figure 20B, respectively. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C) and high (H) IF diets).

For Arpp21, Cldn4, Wap, and Extl1 effects of the high IF diet on mRNA expression was diminished when E2 was administered. The animals did not show a difference in Gadd45 gene expression without E2 exposure but when challenged with E2 for three consecutive days high lifelong IF levels increased the estrogenic response.

Similarly, lifelong exposure to different IF doses, *i.e.*, CC, LL and HH led to significant group wise changes in gene expression (Figure 25). Of the seven regions that were affected by a pre OVX exposure (Figure 23), five regions additionally showed dose dependent effects, *i.e.*, Arpp21, PTEN, Cldn4, Esr1, and Extl. In agreement with the observed U-or bell-shaped trends in methylation changes, relative mRNA expression was affected by lifelong exposure to the low IF dose (group LL) (also see Figure 20).

Comparable to the modifications in DNA methylation, E2 treatment from PND 94 to 97 had nearly no impact on the transcriptional regulation of the analyzed candidate genes. Solely for Extl1, exposure to high IF levels increased the susceptibility to the estrogen responsiveness, but only if administered in a lifelong fashion (Supplemental figure 4).

A summarizing figure for relative mRNA expression across all experimental groups can be found in the supplements (Supplemental figure 6).

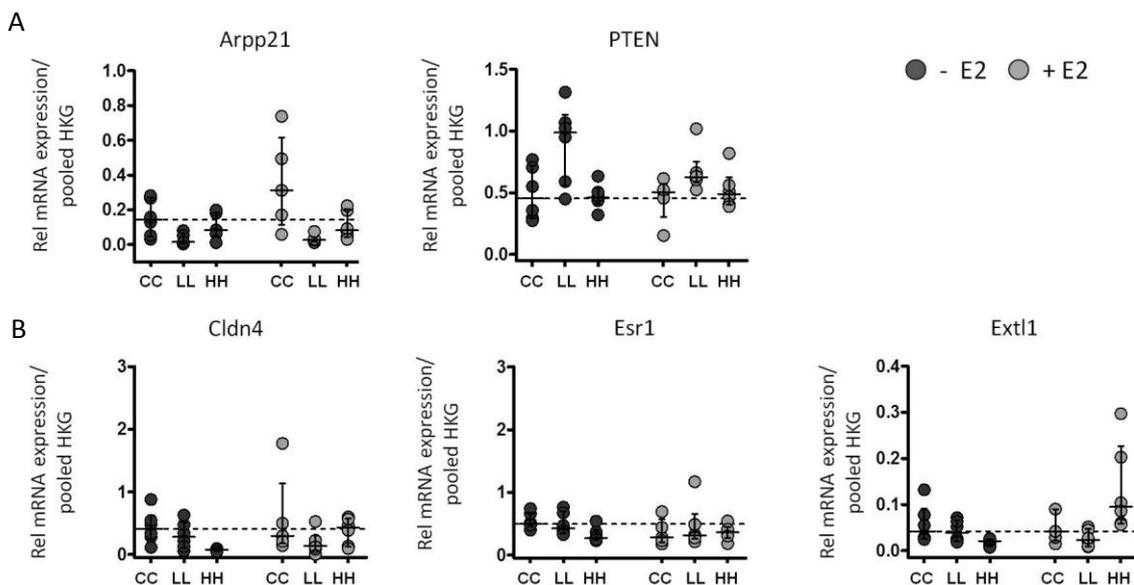


Figure 25: Dose dependent trends of a lifelong IF exposure (LL and HH) on gene expression.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97), summarizing A: bell- or U-shaped and B: linear alterations in gene expression. Median rel. mRNA expression to pooled housekeeping genes (HKG: Hprt, TBP and β -actin) is indicated as horizontal lines, with vertical lines depicting interquartile range. Median rel. expression of the CC group is indicated as horizontal dashed line. For group wise comparison and statistical calculations of dose-trend-relationships see Supplemental figure 3 and Figure 20B, respectively. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), and high (H) IF diets).

4.3 Lifelong exposure to dietary IF affects specific endpoints in ACI rats

In a second study, we focused on the impact of a lifelong dietary IF exposure on the kinetics of DNA methylation changes in MGs of ACI rats during developmentally critical time windows, *e.g.*, embryonic/fetal, early postnatal, and pubertal periods, as well as on the cancer preventive efficacy during E2-driven carcinogenesis of the MG.

4.3.1 Kinetics of DNA methylation in the course of normal MG development

For the lifelong IF intervention protocols, ACI rats were assigned to control C or H diets prior to mating. Female offspring in the H diet group was exposed to IF during all developmental windows of fetal and neonatal, as well as pubertal and adolescent development. In order to better understand the impact of IF on the kinetics of DNA methylation changes during normal MG development, rats were sacrificed before and after puberty at PND 21 and PND 50±1, during adolescence at PND 81±2 and PND 97±2, as well as in adulthood at PND 180±3 (Figure 26).

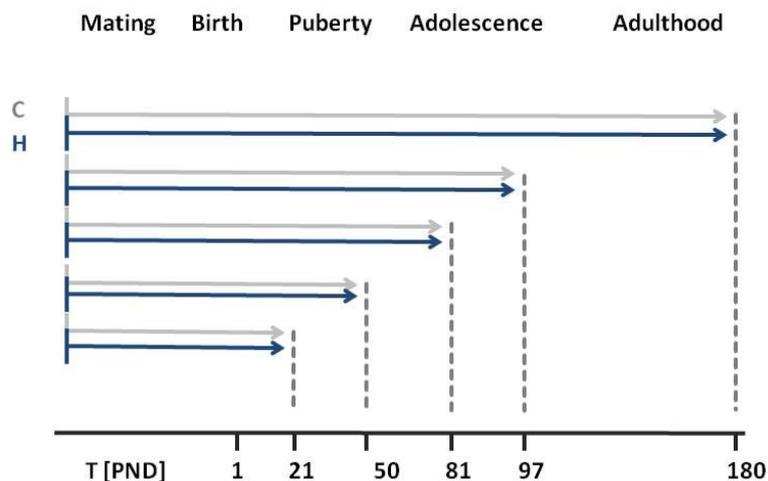


Figure 26: Experimental outline to study DNA methylation kinetics during MG development.

August Copenhagen Irish rats received either control (C) or high IF diet (H), containing 3 ppm and 503 ppm IF aglycone equivalents, respectively, and were sacrificed before and after puberty (PND 21 and 50), during adolescence (PND 81 and 97) as well as in adulthood (PND 180). PND: post natal day

4.3.1.1 Onset of puberty and effects on physiological parameters (Results provided by Möller/Vollmer TU Dresden).

In our previous study in Wistar rats, we demonstrated potential effects of IF on several endocrinological parameters such as onset of puberty and UWW (Figure 9).

For comparison, we also analyze ACI rats for IF-induced effects on puberty onset and UWW. Lifelong IF exposure resulted in an about 1.5 days earlier vaginal opening compared to the control group C

(not significant $p=0.3452$, Figure 27A). Relative UWW normalized to body weight were not affected by dietary exposure to IF (Figure 27B), but we observed significant differences between the age groups ($p=0.0066$) with a pronounced reduction of UWW after puberty at PND 50 and 81 and an increase above prepubertal levels at older age.

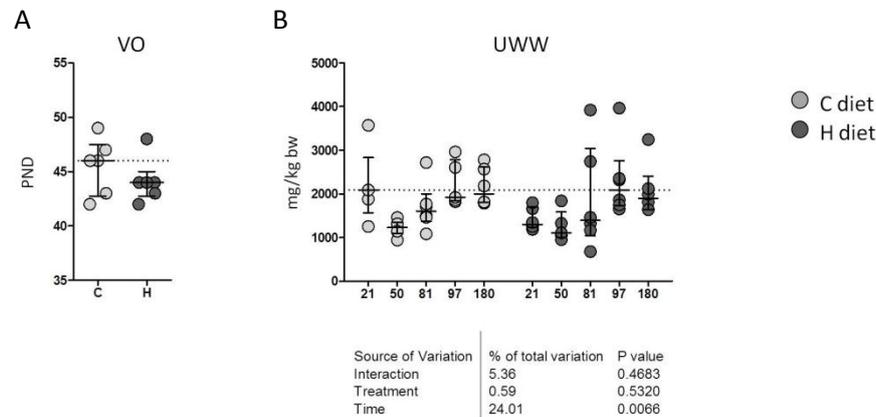


Figure 27: IF-induced effects on vaginal opening and relative uterus wet weight across all developmental windows.

A: Vaginal opening (VO), B: uterine wet weight (UWW) from PND 21-180. Each dot in the plots represents one animal (light grey, control diet and dark grey, IF diet). Horizontal dashed lines indicate median values of the control group (PND 21, C diet), vertical lines indicate interquartile range. Statistical significance of differences was calculated using Two-way ANOVA with the factors: Time (age), Treatment (IF intervention), and interaction between age and IF intervention.

4.3.1.2 Estrogenic and proliferative response of the MG across developmental stages (Results provided by Karin Klimo).

In order to monitor the impact of lifelong IF intervention on MG development, we monitored E2 responsiveness by determining mRNA expression of selected reference genes for proliferation (*e.g.*, Ki67, PCNA) and hormone receptors (*e.g.*, Pgr, Esr1, Esr2), see Figure 28 and Figure 29.

During normal development of the MG relative mRNA expression of the proliferation marker Ki67 increased from PND 21 to PND 180 (borderline significant $p=0.0671$). Interestingly, IF intervention led to a significant increase in PCNA mRNA expression at prepubertal phase ($p=0.0147$) which returned to basal levels observed in control groups from post-puberty onwards. No signal for Pgr mRNA expression was observed at PND 21 in both diet groups. From puberty onwards, Pgr mRNA expression significantly increased, displaying a clear age effect ($p=0.0229$) which was not affected by the dietary IF intervention. However, transcript levels of Esr1 and Esr2 exhibited no significant alterations during MG development.

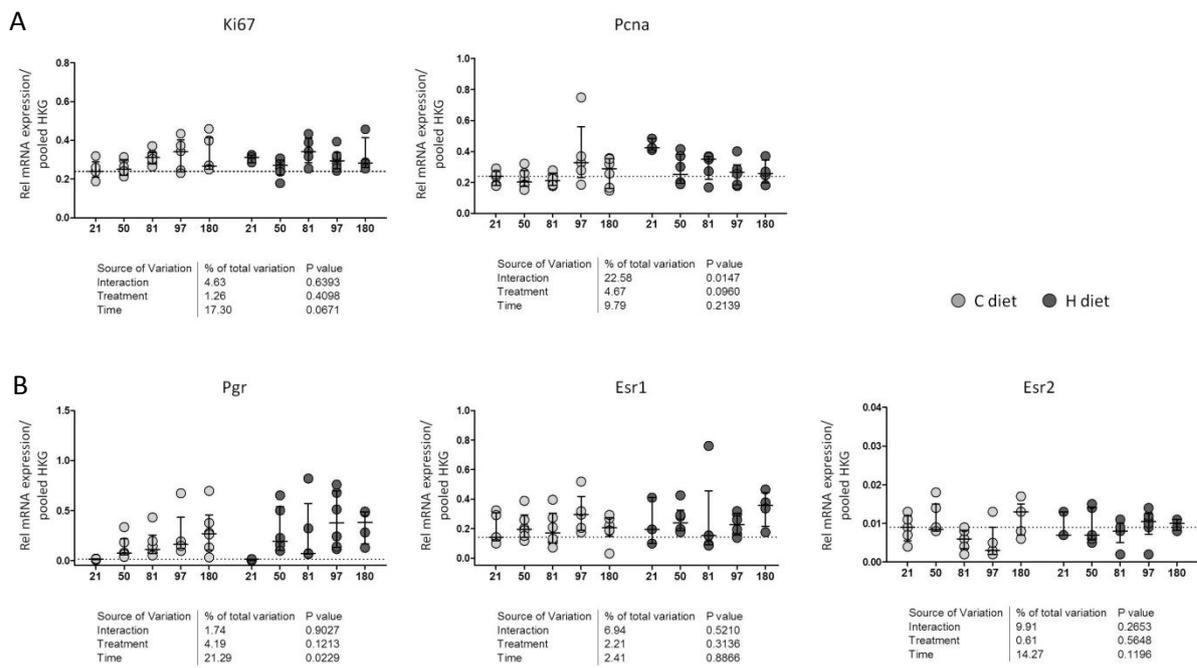


Figure 28: Gene expression of E2-responsive genes across all developmental windows.

Relative Gene expression of A: proliferation markers and B: hormone receptors during all developmental windows from PND 21 until PND 180. Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median methylation values from the C group at PND 21. Vertical lines indicated interquartile range. Statistical significance of differences was calculated using Two-way ANOVA with the factors: Time (age), Treatment (IF intervention), and interaction between age and IF intervention. For comparison between consecutive developmental windows see Figure 29.

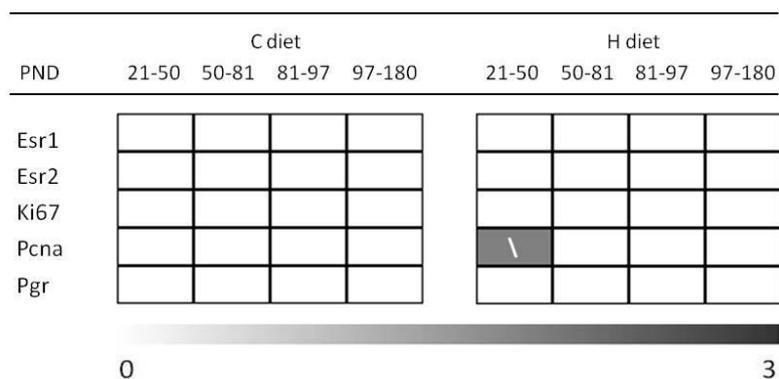


Figure 29: Statistical comparison of gene expression of E2-responsive genes between consecutive developmental windows.

For statistical analysis of pair wise comparison a parametric beta regression model was used to test for significant differences in gene expression. Pairwise comparisons were adjusted for multiple testing (Tukey test). Shapes in boxes correspond to changes in gene expression /: increase, \: decrease between age groups indicated above. Different shades of grey indicate the negative decadal logarithm of three distinct P-values ($p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, ranging from light to dark grey, white = not significant).

4.3.1.3 IHC in mammary gland tissue (Results provided by Möller/Vollmer TU Dresden).

For comparison with mRNA expression, protein levels of Ki67 and PR were determined in all age groups by IHC. The proliferation index (percentage of nuclei stained positive for Ki67) gradually increased during normal development of the MG in the control group C, from 10 % and 20 % at PND 21 and 50, to 37 % at PND 180. Interestingly, lifelong IF intervention increased proliferation indicated by Ki67 protein expression at young age (PND 21 and 50) to levels similar to those seen at older age (on average 46 % positively labeled nuclei) and even above the level detected in adult ACI rats maintained on C diet at PND 180 (Figure 30A).

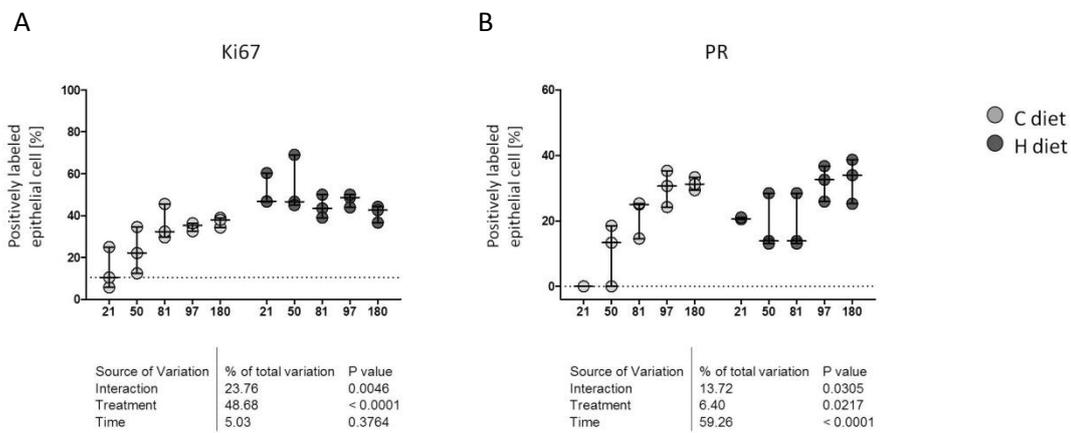


Figure 30: IF-induced effects on Ki67 and PR protein expression during MG development.

A: Expression of the proliferation marker Ki67 and B: the E2-responsive gene PR in the MG determined by IHC. Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). Horizontal dashed lines indicate median expression level of the control group (PND 21, C diet), vertical lines indicate the range. Statistical significance of differences was calculated using Two-way ANOVA with the factors: Time (age), Treatment (IF intervention), and interaction between age and IF intervention.

No signal for PR was observed at PND 21 in the C diet group. From puberty onwards, PR expression in MGs significantly increased to 30 % positively labeled nuclei at PND 180, indicating a clear age influence ($p < 0.0001$). Dietary IF intervention significantly ($p = 0.0217$) increased the percentage of positively labeled nuclei especially in the younger animals (up to 20 % at PND 21). For both, Ki67 and PR, we observed a statistically significant effect of the combination of IF intervention and age.

4.3.2 Implementation of a RRBS library preparation protocol and trouble shooting

Similar to the investigations in Wistar rats, we intended to perform genome-wide methylation analyses to investigate the impact of the IF intervention on DNA methylation in the MG during development. We decided to establish the RRBS technology for the ACI rat experiments, since it avoids the drawbacks of the enrichment-based MCIp method used in the Wistar study. The

advantage of RRBS lies in the fact that this method is not enrichment-based and provides a direct quantitative readout of single nucleotide methylation levels for biostatistics comparison of groups.

In cooperation with Christoph Bock (CeMM, Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna), we have established a RRBS library preparation pipeline. Next-generation sequencing of the libraries was performed at CeMM.

4.3.2.1 RRBS library preparation

Briefly, genomic DNA is digested with *MspI* creating fragments with a 5'-CpG at the terminal ends. Fragments are end repaired, A-tailed, and ligated to standard Illumina adapters. Libraries are multiplexed by six in one library pool and epigenetic methylation marks are converted into genetic marks by bisulfite treatment. This information can be read by genomic sequencing using the sequencing by synthesis technology provided by Illumina.

The *MspI* restriction endonuclease cuts at C↓CGG and is insensitive to CpG methylation. Digestion with *MspI* generates a predictable number of DNA fragments, including highly redundant fragments from microsatellite DNA (Figure 31A).

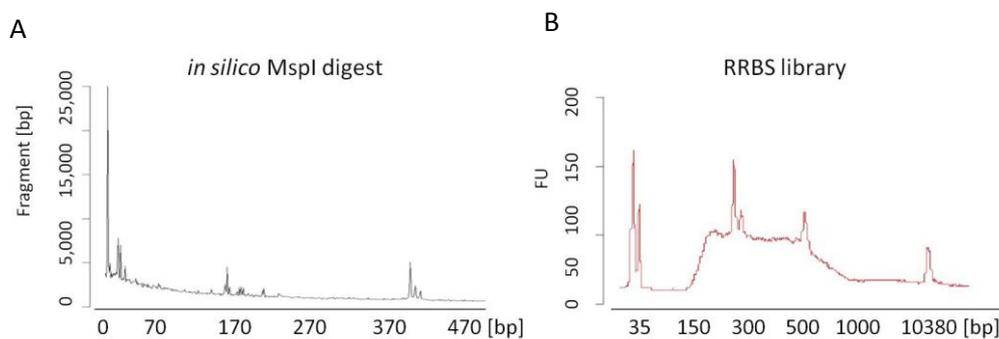


Figure 31: Characteristics of a RRBS library preparation.

A: *In silico* *MspI* digest using the rat genome annotation rn5 and B: Final RRBS library using rat specimen visualized on an Agilent high sensitivity DNA chip. Peaks at 35 and 10,380 bp are internal markers used for DNA quantification as well as adapter dimers. FU: fluorescence units.

Microsatellites are simple sequence repeats of 1 to 4 nucleotide units. These single units are repeated for characteristic number of times, therefore microsatellites can be used as genetic markers, *e.g.*, for genetic fingerprinting in forensics and as positive controls in DNA library preparation. Since all fragments including microsatellite fragments are ligated to Illumina standard adapters, after PCR amplification highly redundant microsatellite DNA is clearly visible in final RRBS libraries, monitored for size confirmation on a Agilent high sensitivity DNA chip. By adding an extra 130 bp (adapter size) to the *MspI* introduced fragment length, we could identify microsatellites at

roughly 300, 330 and 500 bp size/migration time with high fluorescence intensity (FU), corresponding to their redundancy (Figure 31B).

4.3.2.2 Library preparation troubleshooting

RRBS library preparation is multistep process starting from DNA isolation and MspI digestion to End repair, A-tailing, and adapter ligation, multiple clean up steps before and after library multiplexing, as well as bisulfite conversion, and final enrichment PCR.

Potentially each of these steps is prone to errors that jeopardize library integrity.

Loss of fragments

SPRI beads and the amount of corresponding buffers used to remove adapters, dNTPs, and enzymes from the reactions affect the size of fragments eluted from the beads (or that bind in the first place) and determine size distribution of the final library. The combination of amount of beads and buffer is a crucial step of optimization. Particularly, the concentration of polyethylene glycol defines the binding of DNA to the beads. AMPure XP magnetic beads show high affinity to double stranded DNA and 1 μ l is able to retain more than 1 μ g DNA. Since RRBS is a low input technique we used only 100 ng DNA for library preparation. Therefore, SPRI beads are highly overdosed. To avoid overdosing, AMPure XP magnetic beads were purified from the delivered buffer and were diluted 1:5 with a self-made "crowding reagent". Final fragment binding was controlled for pure and diluted SPRI beads showing optimal DNA binding capacity at low DNA input and two different binding regimes (0.75x and 1.5x volumes of beads, Figure 32A and B, respectively). As expected, by increasing the ratio of SPRI:DNA from 0.75x to 1.5x smaller fragments were bound and eluted from the beads. The overall quantity of retained DNA was not affected as detected by identical fluorescence intensities, underscoring the high quality DNA binding ability of the beads diluted by the crowding reagent.

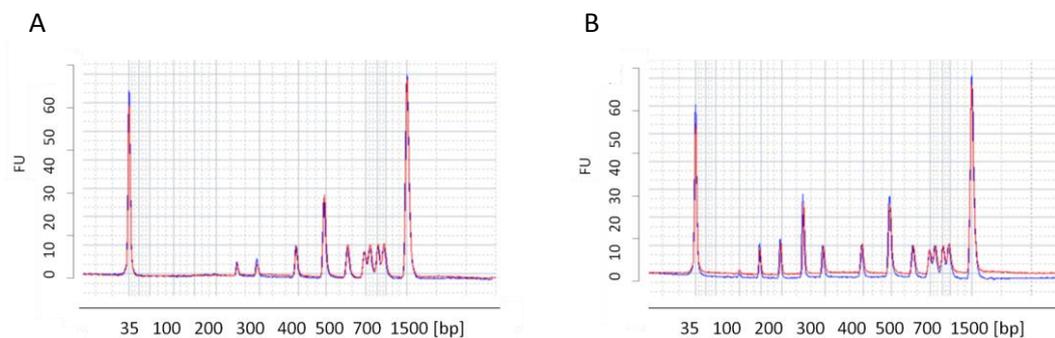


Figure 32: DNA binding capacity of pure and diluted AMPure XP SPRI beads.

Determination of DNA binding of pure (blue) and 1:5 diluted (red) AMPure XP SPRI beads by adding A: 0.75x and B: 1.5x volume of beads to a Fermentas gene ruler. FU: fluorescence units Peaks at 35 and 1500 bp are internal markers used for DNA quantification.

For the epigenetic profiling of the kinetics of DNA methylation during normal developmental of MGs in ACI rats, we prepared four libraries for 24 samples (MGs of six ACI rats on two diets each, at PND 21 and 50). Final enrichment PCRs were cleaned up using pure SPRI beads.

Although pure and diluted beads performed identical in the initial experiment depicted in Figure 32, in this library preparation experiment smaller fragments less than 250 bp were not bound to the beads and got lost in the final library (Figure 33A). The fragments smaller than 250 bp were not retained in the four libraries of the 24 MG samples, with a slightly higher amplification of fragments with sizes between 300 bp to 500 bp in the pink library compared to the three other libraries.

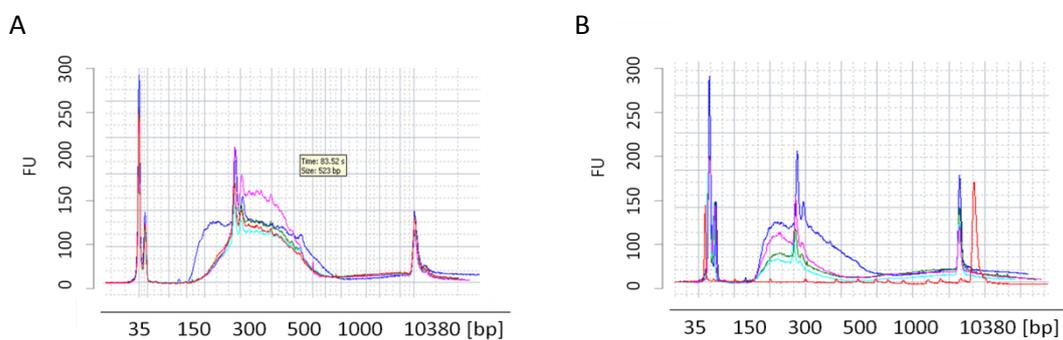


Figure 33: Troubleshooting: Loss of fragments during RRBS library preparation.

Final library quality was monitored on Agilent high sensitivity DNA chip, detecting fluorescence units (FU) vs. size/migration time. Electropherograms showing loss of fragments with A: sizes less than 250 bp (compare representative library profiles with the blue profile of an optimal library) and B: a general loss of amplified DNA and larger than 400 bp (compare profiles to the optimal profile in A and decreasing FU per RRBS library). Peaks at 35 and 10,380 bp are internal markers used for DNA quantification as well as adapter dimers.

Fragment loss also occurred when specific thermo cyclers were used for the enrichment PCR (examples in Figure 33B). Here, particularly longer fragments over 400 bp or even the overall amounts of libraries were reduced. Varying ramping rates even of identical cyclers should be taken into consideration for library preparation, especially if protocols do not perform well in other hands. Therefore, besides fresh and clean consumables, the choice of the cycling machine and the corresponding ramping rates should be controlled properly.

Bisulfite conversion rate

In order to maintain methylation information during PCR amplification, DNA samples are subjected to bisulfite conversion. Bisulfite treatment of DNA leads to a chemical deamination of unmethylated cytosines (C) to uracils (U). In subsequent PCR reactions, all uracils (from unmethylated Cs) are amplified as thymines, allowing discrimination of unmethylated and methylated Cs at single CpG resolution, since only methylated Cs are amplified as cytosines. Achieving a consistent C to U

conversion of unmethylated DNA is crucial for obtaining reliable quantitative methylation data. Since several bisulfite conversion protocols were available in our lab, RRBS libraries were bisulfite treated following three different protocols to determine the procedure leading to highest bisulfite conversion rates.

The Methylation Direct Kit from Zymo Research was used with two different cycling conditions, varying in the time the DNA is heated to 95 °C. Protocol 1 (30 s): 16x (30s at 95°C, 10 min at 60°C) and Protocol 2 (60 s): 16x (60s at 95°C, 10 min at 60°C). For a third approach, the standard Zymo Research methylation Kit was used according to manufacturer's recommendations; Protocol 3: 16x (15s at 95°C, 60 min at 50°C). This protocol runs overnight with longer incubation at a lower temperature compared to Protocols 1 and 2.

In general, all three methods considerably harm the DNA, leading to an overall library loss of >90 % as detected by qPCR (data not shown). However, different bisulfite conversion protocols also affect the loss of fragments. Particularly, conversion protocols 2 and 3 reduced the overall amount of library which can be seen by lower fluorescence intensities using the Agilent high sensitivity DNA chip (Figure 34A). However, protocols 2 and 3 showed the most reliable bisulfite treatment with conversion rates between 98.5-99 % (Figure 34B). For all the RRBS libraries we used bisulfite conversion Protocol 2.

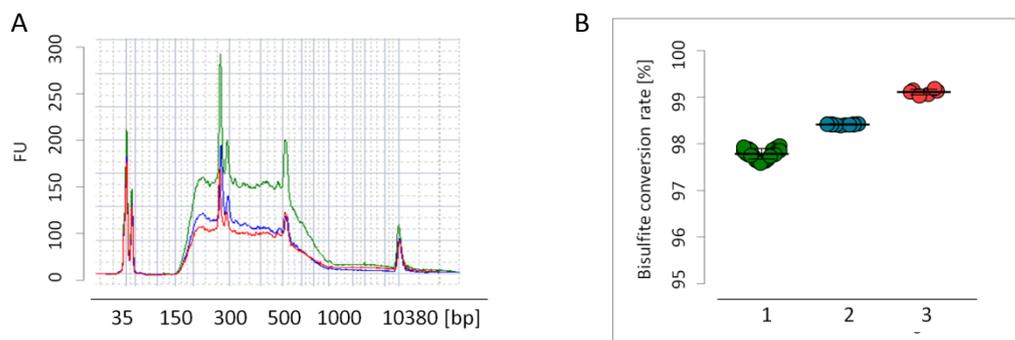


Figure 34: Troubleshooting: Bisulfite conversion rate during RRBS library preparation.

Three different bisulfite conversion protocols of commercially available kits regularly used in our lab were tested for A: loss of library fragments monitored on an Agilent high sensitivity DNA chip and B: bisulfite conversion rates calculated based on cytosine conversions in a non CpG context. Methylation Direct Kit from Zymo Research was used for Protocol 1: 16x (30s at 95°C, 10 min at 60°C) in green and Protocol 2: 16x (60s at 95°C, 10 min at 60°C) in blue. For Protocol 3: 16x (15s at 95°C, 60 min at 50°C) in red, the standard Zymo Research methylation Kit was used according to manufacturer's recommendations. FU: fluorescence units

4.3.3 Epigenetic profiling of developmental effects during puberty

As outlined above, we subjected 24 MG samples to RRBS library preparation and NGS, in close cooperation with Christoph Bock (CeMM, Vienna). After pipeline adjustments for the alignment to

the rn5 rat genome, initial bioinformatics analyses of the sequencing data were also performed at CeMM.

Out of 24.7 M CpGs that are present in the rat genome we covered roughly 2.18 M CpG (unique aligned motifs) with a general mean coverage of 32x per site. By sequencing our libraries on one lane of a Illumina HiSeq2000 machine, we obtained 14-35 M high quality reads per sample with an alignment rate ranging from 73–76 %. Calculations with a bisulfite conversion rate of 98.5% yielded to a mean global methylation of 64 % at PND 21 and 50. A general summary of the obtained sequencing results of pre- and postpubertal ACI rats can be found in Supplemental table 3.

For the identification of DMS, the RnBeads pipeline was used. For efficient detection of DMS resulting from exposure to IF-depleted or -enriched chow within each age group, only CpG sites were considered with a minimum coverage of 20 reads in at least 5 of the 6 samples per dietary group, reducing the number of sites to roughly 165,000 and 168,000 DMS, respectively (Figure 35).

24.7 M CpGs				
Coverage (all)	2,18 M CpGs			
Diet	C	H	C	H
Age [PND]	21 ± 1		50 ± 2	
minCoverage20 (10 out of 12)	165,796		168,024	
	↓		↓	
comparison groups	C vs. H		C vs. H	
minCoverage >10 (12 out of 12), p<0.05	7,607 CpGs		4,646 CpGs	
Mean methylation difference > 5%	5,622 CpGs		3,793 CpGs	
Distance between CpGs < 300 bp	512 CpGs		376 CpGs	
	↓		↓	
genomic position	Promoter		Promoter	
Distance to next TSS [2 kb to -0.5 kb]	267 genes		207 genes	
Candidate genes	Zfp507, Fxyd2, Arpp21, Mmaa, Rpm42, Bpgm, Tlr3, Nr1h3		Niacr1, Shroom2	

Figure 35: Flowchart for the selection of candidate regions for RRBS confirmation from the puberty study.

PND: post natal day, TSS: transcriptional start site,

Further filtering of the dataset for significant ($p < 0.05$) methylation changes of $> 5\%$, less than 10,000 DMS were retained within each age group. Around 7 % of these selected sites were located in regulatory regions (promoters, CpG islands or 5' UTRs). Only 282 DMS were overlapping between both age groups, indicating that within each age group a specific set of CpG sites is significantly affected by IF exposure (Figure 36).

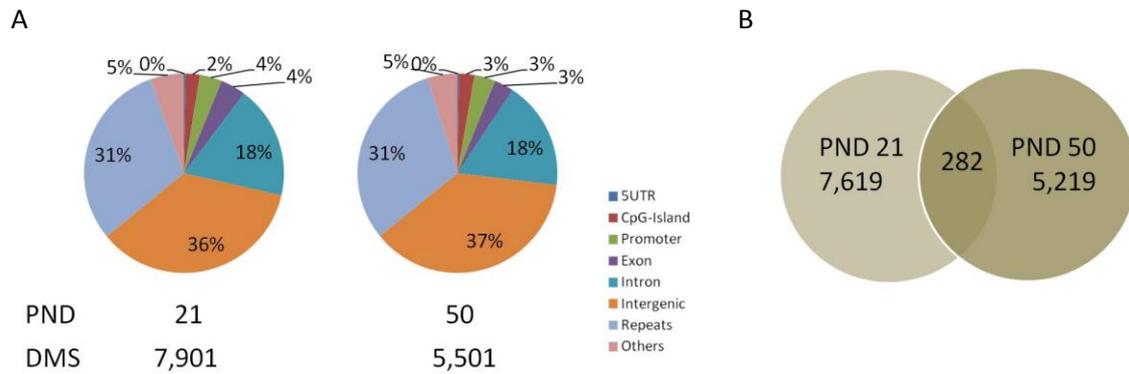


Figure 36: Genomic distribution and overlap of DMS from the puberty study.

CpGs with a minimum coverage of 20 reads in at least 5 of the 6 samples within each age groups were considered and filtered for IF-induced differentially methylated sites (DMS) with significant methylation difference $>5\%$. A: Genomic distribution of DMS and (B) Venn diagram depicting the overlap of DMS between age groups. CpG sites were annotated to 8 categories (indicated by colors) according to the nearest transcription start site and grouped in % relative to the total number of DMS below each chart. PND: post natal day, DMS: differentially methylated site

4.3.3.1 Technical confirmation of quantitative methylation analyses

Although RRBS provides a direct quantitative read out of single nucleotide methylation levels, the validity of the results has to be confirmed by independent methods. Using the mass spectrometry-based EpiTYPER MassARRAY technology we are able to detect methylation differences $\geq 5\%$ by quantifying median methylation of CpG units in a high throughput manner.

Selection of candidate regions for RRBS confirmation

Candidate site selection from the RRBS data was based on a minimum coverage of 20 reads in at least 5 of the 6 samples, with a minimum coverage of 10x for each sample. CpGs selected for validation showed a significant methylation difference of $>5\%$ between groups, with two or more sites in close proximity <300 bp distance. Since the relationship between DNA methylation and gene expression is best understood for promoter regions, we additional filtered our data for DMS located between 2 kb upstream and 0.5 kb downstream of a TSS. After applying these settings, IF-induced differential methylation was identified for 512 CpGs in 267 gene promoters in MGs of prepubertal and for 376 CpGs in 207 gene promoters in MGs of postpubertal ACI rats.

Ten candidate regions/gene promoters were selected for MassARRAY validation (Figure 35). The respective proteins are associated with energy homeostasis and metabolism (Fxyd2, Bpgm, Mmaa and Niacr1), cell structure (Shroom2), have been shown to be downregulated during breast carcinogenesis (Arrp21) or directly induced tumor growth and stem cell-like phenotype in xenografts (Tlr3). Zfp507 is a not further characterized zinc-finger protein, and has not yet been investigated with respect to tumor formation or MG development. Rpm42 might affect cell cycle control by

binding to p21, and Nr1h3 is involved in homeostasis of cholesterol and metabolism of steroid hormones. A summary of all RRBS candidates and EpiTYPER amplicon information can be found in Supplemental table 4.

RRBS confirmation using MassARRAY technology

Primers for the MassARRAY validation of ten candidate regions were designed and conditions for PCR amplification were optimized. Each DMR was represented by at least 2 single CpGs or a CpG unit consisting of 2 CpGs.

The majority of the selected regions were identified in the PND 21 group, except for Shroom2 and Niacr1 which showed differential methylation in animals at PND 50. Niacr1, Zfp507, Fxyd2, Arpp21, and Mmaa showed high methylation levels around 1 in RRBS that were reduced by intervention with IF (Figure 37). For Niacr1, we observed significant methylation differences of 19-24 % by RRBS between the dietary groups at PND 50, but not at PND 21. Analyzing the same region with MassARRAY technology we could not confirm this IF-induced methylation change. Similarly, for Zfp507, Fxyd2, Arpp21, and Mmaa RRBS identified DMS between the C and H group at PND 21 that we were not able to confirm by the EpiTYPER technology. However, with the exception of Zfp507 methylation in the PND 50 group (RRBS had indicated unmethylated CpG sites), we were able to validate at least the overall degree of methylation of the selected regions above 50 % absolute methylation.

For Zfp507, but also in general, group wise methylation levels were much more homogeneous when determined by MassARRAY in comparison to the RRBS data. This indicates that the methylation differences detected by RRBS were at least partly due to technical artifacts. Interestingly, for Arpp21, Niacr1, and Fxyd2 we detected by MassARRAY pronounced methylation differences between the individual age groups with an overall increase of 25 % in the PND 50 group for Arpp21 and a decrease of 10-12 % for Niacr1 and Fxyd2. Since we had access to MG tissue of various age groups (Figure 26), we further followed up these observations in section 4.3.3.2.

For Rpm42, Bpgm, Tlr3, and Nr1h3, we focused on PND 21 only. With RRBS, the selected regions for Tlr3, Nr1h3, and Bpgm showed low levels of methylation in the control group, and IF-intervention led to a significant gain (Figure 38). Rpm42 methylation levels were in the range of 60-90 % in the control group and also were hypermethylated in the H diet group. Differential methylation detected by RRBS between C and H diet groups could not be confirmed by MassARRAY, although the level of overall methylation was consistent, with the exception of Rpm42.

For Shroom2, individual methylation values for the CpG sites covered by RRBS reads were not obtainable by MassARRAY, therefore a direct comparison was not possible.

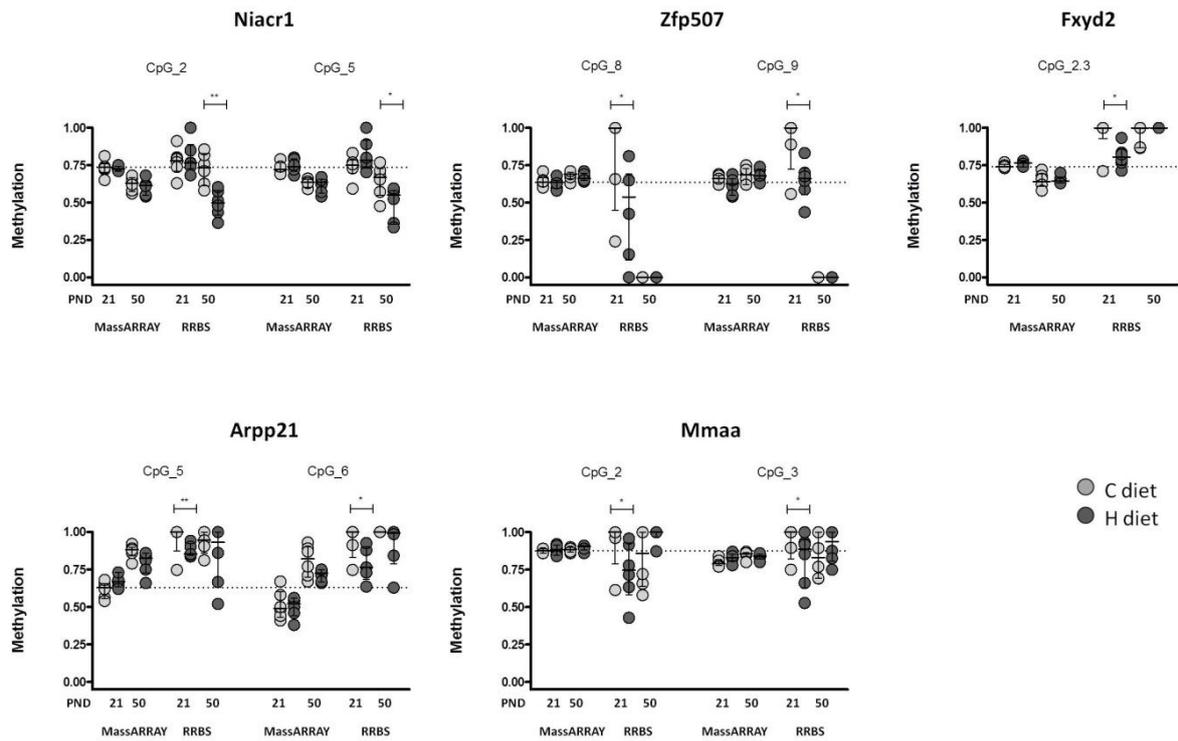


Figure 37: Technical confirmation of highly methylated DMS by MassARRAY technology from the puberty study.

Absolute methylation of the DMS analyzed by MassARRAY (left) and RRBS (right). Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). Horizontal dashed lines indicate median methylation values of the first CpG (PND 21, C diet), vertical lines indicate interquartile range. PND: post natal day.

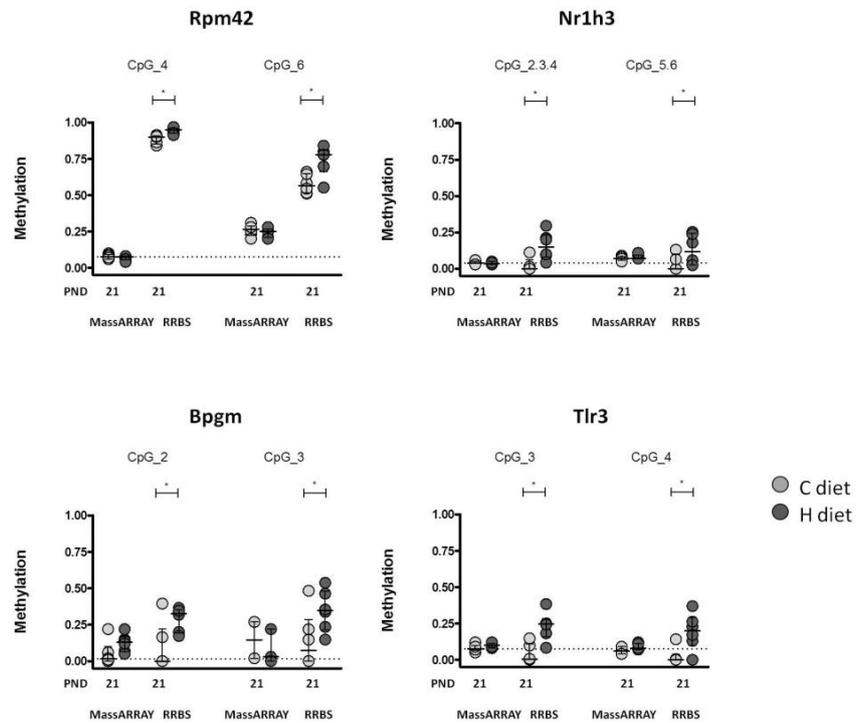


Figure 38: Technical confirmation of lowly methylated DMS by MassARRAY technology from the puberty study.

Absolute methylation of the DMS analyzed by MassARRAY (left) and RRBS (right). Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). Horizontal dashed lines indicate median methylation values of the first CpG (PND 21, C diet), vertical lines indicate interquartile range. PND: post natal day.

Although we were not able to confirm IF-induced methylation changes by MassARRAY, our results indicated that the general methylation levels detected by both technologies were comparable. We performed a correlation analysis considering methylation data for the CpGs in the MassARRAY amplicon that covered the DMS identified by RRBS. Methylation was measured in a zero-to-one scale, and for each of the candidate regions, median methylation was calculated for the selected CpGs (Figure 39). In total, 166 values for 24 animals were plotted against each other and showed highly significant correlation (Spearman correlation coefficient ρ : 0.7038, 95% CI = 0.6149-0.7751 and $p < 0.0001$).

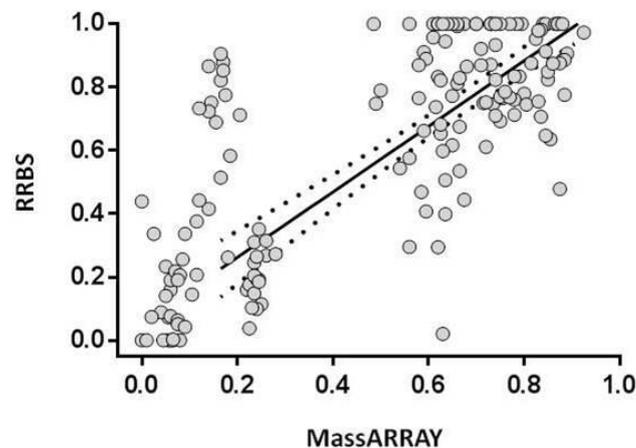


Figure 39: Correlation of quantitative RRBS and MassARRAY data from the puberty study.

Each dot in the correlation plots represents one median methylation value per animal obtained for 24 animals in nine candidate regions (Spearman correlation coefficient ρ : 0.7038, 95% CI = 0.6149-0.7751 and $p < 0.0001$).

4.3.3.2 Quantitative analysis of candidate regions across all developmental windows

Since we had access to MG tissue of various age groups (Figure 26), we were able to follow up age-induced changes in DNA methylation of *Niacr1*, *Fxyd2*, *Arpp21*, *Mmaa*, *Zfp507*, and *Shroom2* from early development until adolescence and adulthood (PND 21-180). Additionally, we investigated *Esr1* and *Extl1* promoter methylation during MG development, two interesting candidate genes obtained from the MClp-Seq approach in Wistar rats (Figure 40).

For *Niacr1* and *Fxyd2*, methylation levels gradually decreased with advancing age by about 20-30 % in comparison with the prepubertal rats at PND 21 (Figure 40A). Temporarily at PND 97, methylation of *Niacr1* was significantly higher in the H diet when compared to PND 80 or 180 (see Figure 41). For *Arpp21*, *Mmaa*, and *Extl1* methylation levels increased with age (Figure 40B). Methylation of *Extl1* increased up to PND 97 but decreased in older animals to levels below those of prepubertal rats at PND 21 (significant for C diet). Notably, lifelong H diet prevented the loss of DNA methylation

between PND 97-180. Methylation in the promoter regions of *Zfp507*, *Shroom2* and *Esr1* also gradually increased significantly by 5-10 % starting at PND 21 or 50 (Figure 40C).

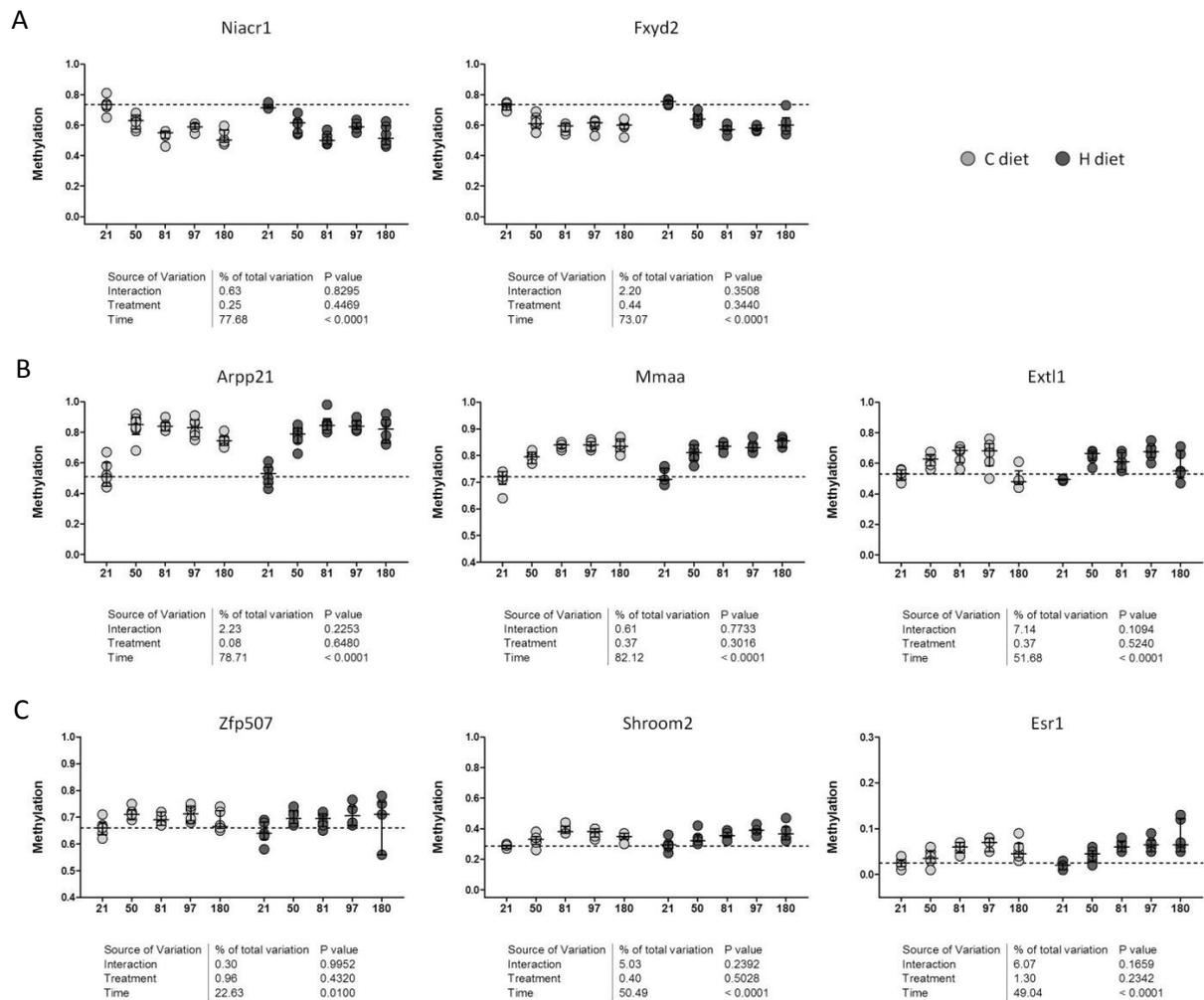


Figure 40: Quantitative analysis of candidate regions across all developmental windows.

Median methylation of selected candidate regions in MGs between PND 21-180, with focus on DMS with confirmed differential methylation levels between PND 21 and 50 (see Figure 37). Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, IF diet). The horizontal dashed line indicates median methylation values from the C group at PND 21. Vertical lines indicate interquartile range. For statistical analysis, Two-way ANOVA was used with the factors: Time (age), Treatment (F intervention), and interaction between age and IF intervention. For comparison between consecutive developmental windows see Figure 41.

In summary, MassARRAY analysis revealed significant age-dependent effects on DNA methylation representing the MG methylome before and after puberty (PND 21 and 50), as well as in adolescence, and adulthood (PND 81-180) with most pronounced effects during puberty (PND 21-50, Figure 41). IF exposure within one age group led to methylation changes of about 5 % which however, did not reach statistical significance.

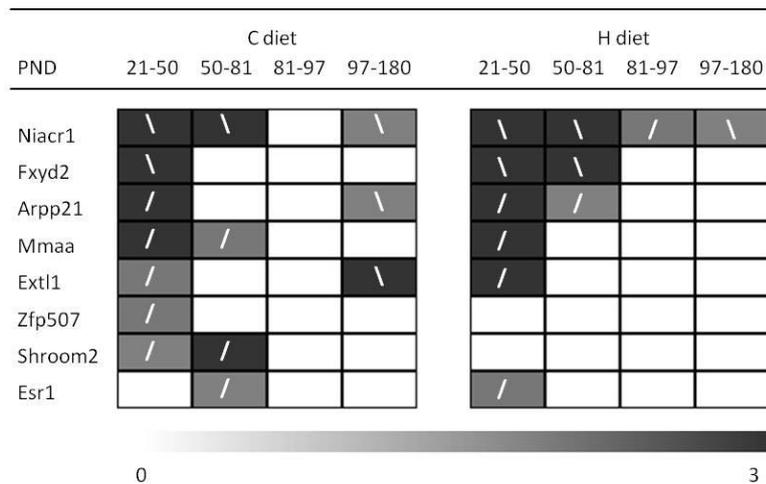


Figure 41: Statistical comparison of candidate regions between consecutive developmental windows.

For statistical analysis of pair wise comparison a parametric beta regression model was used to test for significant differences in DNA methylation. Pairwise comparisons were adjusted for multiple testing (Tukey test). Shapes in boxes correspond to changes in DNA methylation /: increase, \: decrease between age groups indicated above. Different shades of grey indicate the negative decadal logarithm of three distinct P-values ($p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, ranging from light to dark grey, white = not significant).

4.3.4 Gene expression analyses (performed by Karin Klimo)

4.3.4.1 Candidate gene expression and correlation with DNA methylation changes

In order to determine whether changes in DNA methylation might be correlated with changes in gene expression, RT-qPCR was performed. Gene expression levels were measured for Niacr1, Fxyd2, Arpp21, Extl1, Zfp507, and Esr1. No gene expression data was available for Mmaa and Shroom2 due to unsuccessful primer design. We were not able to identify any significant age- or diet-induced gene expression changes in the candidate regions obtained by RRBS (Figure 42).

Consequently, correlation analyses for the comparison of promoter methylation and gene expression did not reveal significant correlation. Spearman correlation coefficient rho was in the range of -0.01 to -0.04 for negative and 0.02 to 0.23 for positive correlation (data not shown).

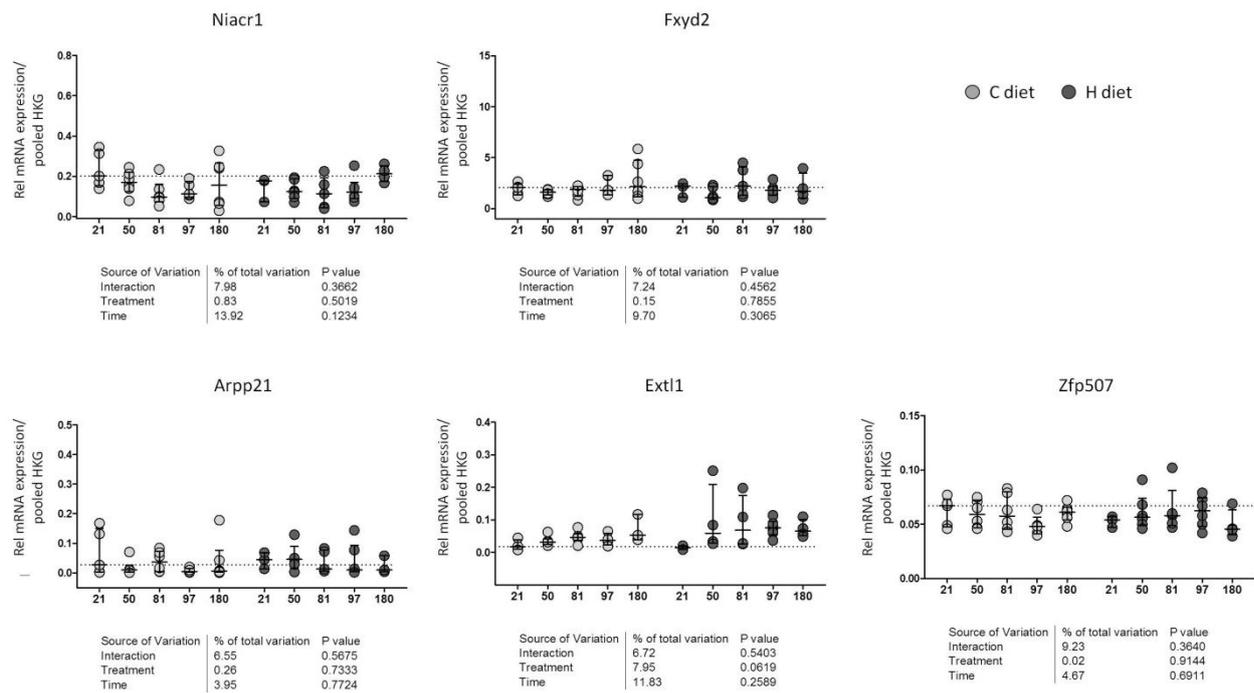


Figure 42: Gene expression of the candidate regions across all developmental windows.

Relative gene expression of selected candidate regions in MGs between PND 21-180. Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median gene expression values of the C group at PND 21. Vertical lines indicate interquartile range. For statistical analysis, Two-way ANOVA was used with the factors: Time (age), Treatment (IF intervention), and interaction between age and IF intervention.

4.3.4.2 Gene expression-based pre characterization using cell type marker genes

Since normal MG development involves a massive reorganization regarding the cytoarchitecture, we selected marker genes for epithelial cells (Keratin 8 (Krt8) and E-cadherin (Cdh1)) as well as for myoepithelial cells (Keratin 5 (Krt5), smooth muscle actin (SMA), and Tp63) to determine alterations in the glandular structure and the cellular composition of the MGs (Figure 43).

We observed significantly increased transcript levels of both myoepithelial and epithelial cell type markers from PND 50 onwards (not significant for Krt5). During MG development strongest up-regulation of cell marker mRNA levels were observed during the pubertal phase between PND 21 and 50 (Figure 44). Interestingly, there was a global effect of the lifelong IF intervention on Krt8 mRNA expression ($p=0.0018$) especially in the older rats. Here, Krt8 transcript levels were significantly higher than those of animals on the C diet.

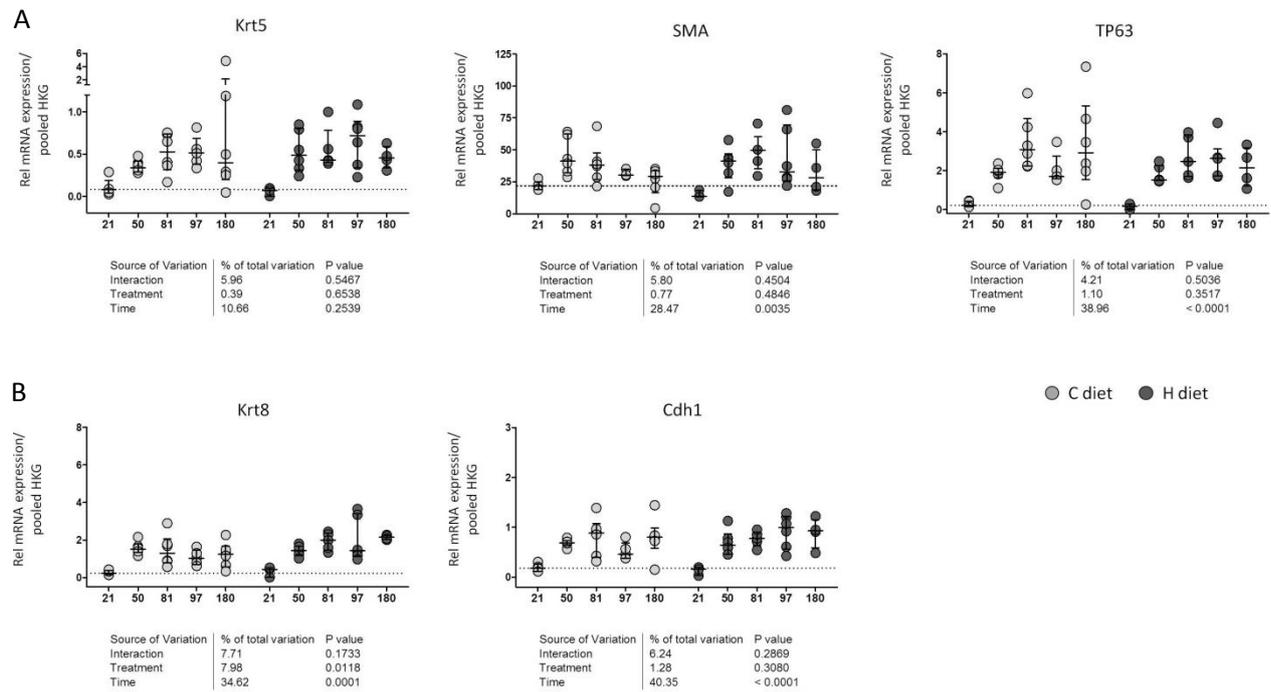


Figure 43: Gene expression of marker genes for epithelial cells across all developmental windows.

Relative gene expression for A: myoepithelial and B: epithelial cell type markers during all developmental windows from PND 21-180. Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates gene expression values from the C group at PND 21. Vertical lines indicated interquartile range. For statistical analysis, Two-way ANOVA was used with the factors: Time (age), Treatment (IF intervention), and interaction between age and IF intervention. For comparison between consecutive developmental windows see Figure 44.

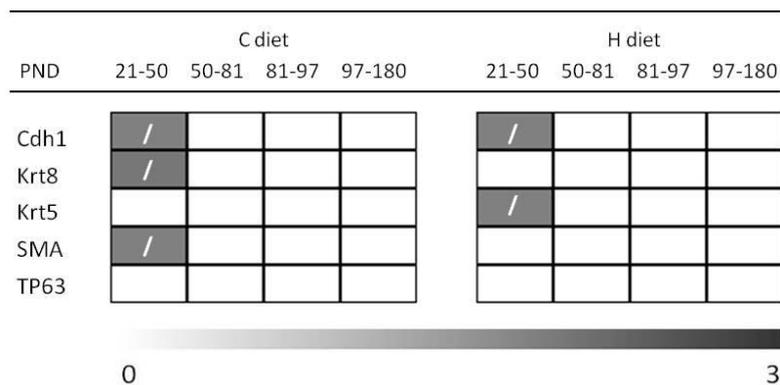


Figure 44: Statistical comparison of gene expression of cell type marker genes between consecutive developmental windows.

For pair wise comparison a parametric linear regression model was used to test for statistically significant differences in gene expression. Pairwise comparisons were adjusted for multiple testing (Tukey test). Shapes in boxes correspond to changes in gene expression /: increase, \: decrease between age groups indicated above. Different shades of grey indicate the negative decadal logarithm of three P-values ($p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, ranging from light to dark grey, white = not significant).

4.3.4.3 IF affect gene expression of key epigenetic enzymes

In order to identify the underlying mechanism of IF and age-induced changes in DNA methylation and gene expression during MG development we measured relative mRNA expression of selected epigenetic modifying enzymes involved in writing and erasing DNA methylation (*e.g.*, DNMT1, 3a, 3b and Tet1).

The relative mRNA expression of DNA methylation modifying enzymes showed a diverse pattern (Figure 45). DNMT1 transcript levels increase in the control group with increasing age, whereas expression showed a bell-shaped pattern in the H diet group with a maximum at PND 81 (not statistically significant). Within the control diet groups, DNMT3a and 3b showed a tendency of lower expression at older age. Lifelong IF intervention reversed this tendency, with lower expression levels at PND 21 in the H compared to the C group and a steady increase in expression up to PND 180, especially for DNMT3b. Tet1 mRNA expression was clearly reduced in the course of MG development in both dietary groups.

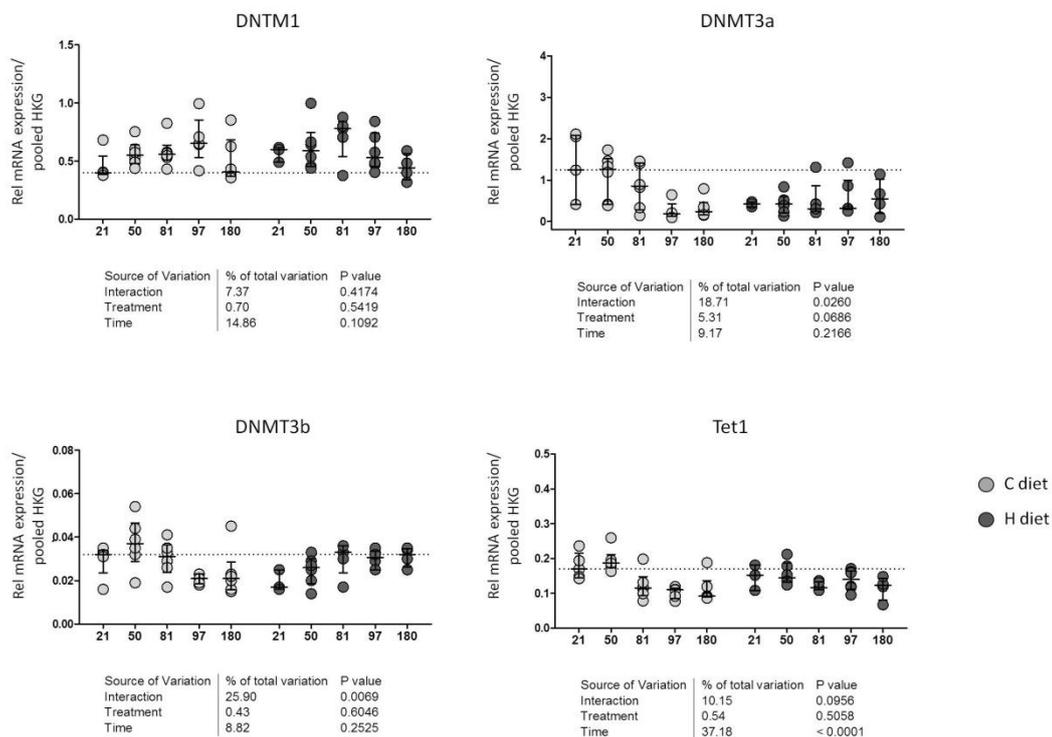


Figure 45: Gene expression of epigenetic key enzymes across all developmental windows.

Relative gene expression of DNA methylation modifying enzymes during all developmental windows from PND 21-180. Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates gene expression values from the C group at PND 21. Vertical lines indicated interquartile range. For statistical analysis, Two-way ANOVA was used with the factors: Time (age), Treatment (IF intervention), and interaction between age and IF intervention. For comparison between consecutive developmental windows see Figure 46.

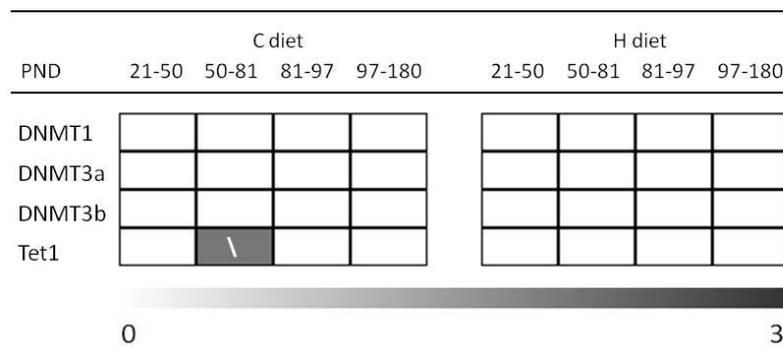


Figure 46: Statistical comparison of gene expression of epigenetic key enzymes between consecutive developmental windows.

For statistical analysis of pair wise comparison a parametric linear regression model was used to test for significant differences in gene expression. Pairwise comparisons were adjusted for multiple testing (Tukey test). Shapes in boxes correspond to changes in gene expression /: increase, \: decrease between age groups indicated above. Different shades of grey indicate the negative decadal logarithm of three P-values ($p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, ranging from light to dark grey, white = not significant).

4.3.5 Epigenetic profiling during E2-induced breast carcinogenesis in ACI rats

The ACI rat strain is highly susceptible to develop mammary tumors upon prolonged exposure to pregnancy levels of exogenous E2 [161]. Six Emca loci determine the breast cancer susceptibility which in part have been identified to be orthologous to loci in human. Further, they have been associated with various risk markers such as mammographic breast density which emphasizes the relevance of the animal model for human breast cancer [166].

4.3.5.1 E2-induced carcinogenesis in the MG (Results provided by Möller/Vollmer TU Dresden)

In order to monitor the impact of IF on MG carcinogenesis, female ACI rats were assigned to control C or H diets prior to mating and female offspring was exposed to IF during all developmental windows of fetal and neonatal, as well as pubertal, and adolescent development (Figure 47). Morphological analyses indicated the highest numbers of highly proliferative and susceptible structures before PND 50 (TEBs, data not shown), therefore rats were implanted with an E2-containing silicon tube to initiate the tumorigenic process at PND 45. In order to further promote the carcinogenic process, the silicon tubes were renewed at PND 175. Previous studies showed reduced numbers of TEBs at PND 50 after exposure to the H diet. Since many processes during early development are regulated by epigenetic mechanisms these processes might be targeted by IF, affecting normal cell growth and susceptibility to breast cancer [144] (reviewed in [2]). For detailed information on animal experimentation see section 3.2.4 as well as Möller et al. 2016 (in revision).

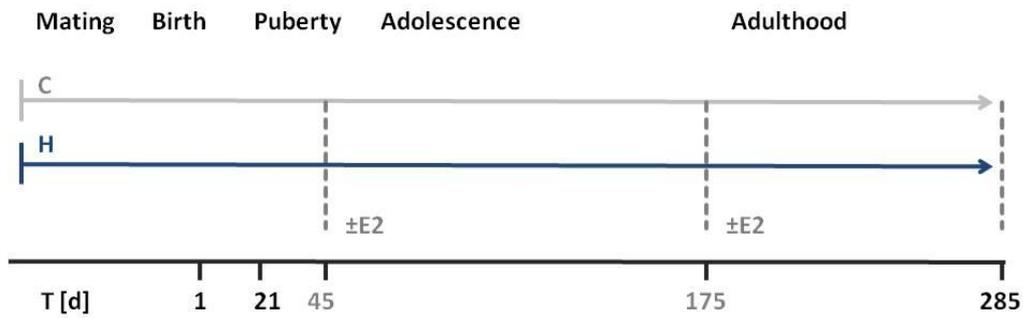


Figure 47: Experimental outline of E2-induced MG carcinogenesis in ACI rats.

August Copenhagen Irish (ACI) rats received either control (C) or high IF diet (H) containing 3 ppm and 503 ppm IF aglycone equivalents. For initiation and promotion of the carcinogenic process, E2-releasing tubes were implanted at PND 45 and PND 175 in half of the animals. All animals were sacrificed at PND 285 to monitor tumor development.

Mammary tumors arose in ACI rats which were implanted with E2-releasing silicon tubes with a median tumor multiplicity of 3.5 tumors per animals. Earliest palpable tumors were detected at postnatal week 25 and 20 for C and H diet, respectively ($p < 0.0001$, data not shown). Although H diet showed a significantly shorter time-to-tumor appearance, tumor multiplicity was reduced by 56 % compared to controls ($p = 0.018$). However, tumors that developed in animals exposed lifelong to H diet grew to larger size with a median total tumor volume of 19.3 cm^3 and wet weights of about 2.5-fold greater than in control animals (Figure 48).

Taken together, these results indicate a chemopreventive effect of lifelong IF exposure on tumor incidence and multiplicity in ACI rats. However, if tumors escaped the preventive effect, they grew to larger sizes (in three out of the six animals), resulting in a higher tumor burden.

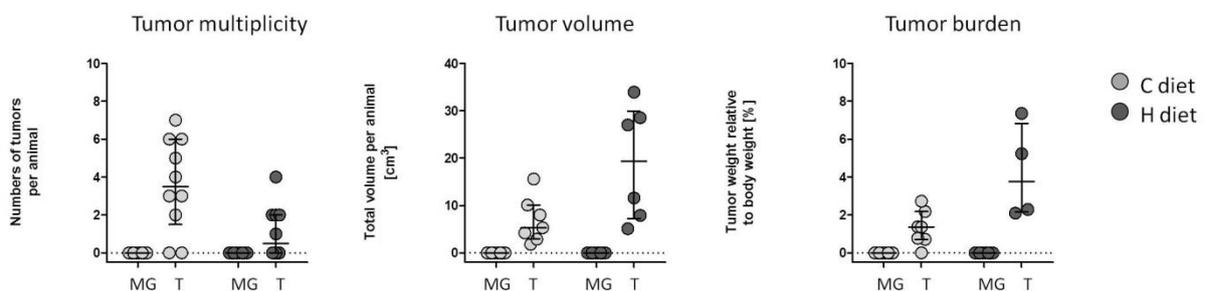


Figure 48: Effects of a high IF diet on tumor characteristics of ACI rats.

Each dot in the scatter plots represents one animal (light grey: C diet, dark grey: H diet). Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. The horizontal dashed line indicates median values from MGs of the C group. For statistical analysis, MG: mammary gland, T: tumor

IF treatment affects various physiological parameters (Results provided by Möller/Vollmer TU Dresden)

In order to assess the potential effects of lifelong IF intervention on endocrinological parameters, we determined UWW and measured mRNA as well as protein expression of prototypical estrogen responsive genes and classical proliferation markers during E2-induced breast carcinogenesis.

Interestingly, UWW was slightly increased at PND 285 by 15 % in both groups on H diet compared to control diet groups, suggesting an estrogenic effect of IF exposure (borderline significant $p=0.0508$).

For Ki67 protein staining determined by IHC, E2 treatment significantly increased the proliferation index from 0 % to 11 % positively labeled nuclei in MGs of E2-exposed animals not bearing frank tumors (MG/T) ($p=0.0028$). Although IF intervention showed a pro-estrogenic effect in the uterus, no significant impact of the lifelong IF diet was observed in healthy and neoplastic MGs in the H group.

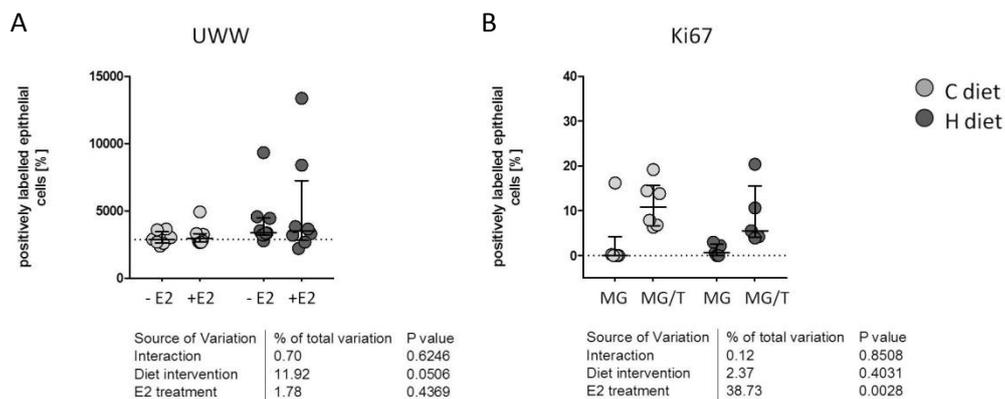


Figure 49: Uterus wet weight and Ki67 protein expression in MG and tumors.

A: Uterus wet weight (UWW) and B: IHC of the proliferation marker Ki67 in response to E2 stimulation. Each dot in the scatter plots represents one animal (light grey: C diet, dark grey: H diet). Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. The horizontal dashed line indicates median values from the C group without estrogen (E2) stimulation. For statistical analysis, Two-way ANOVA was used with the factors: Diet intervention, E2 treatment, and interaction between diet intervention and E2 treatment. MG: mammary gland, MG/T: MG without detectable tumors

Impact of IF intervention on estrogenic and proliferative response in MGs and E2-induced tumors

In order to monitor the impact of lifelong IF intervention during E2-induced MG carcinogenesis, we determined mRNA expression of selected reference genes for proliferation (e.g., Ki67, PCNA) and hormone receptors (e.g., Pgr, Esr1, Esr2) in MGs of untreated animals, as well as in MG and tumor tissue of E2-exposed animals (MG/T and T, respectively). Transcripts levels of the proliferation markers Ki67 and PCNA increased in tumors independent of the dietary regime and confirmed the observations made by IHC (borderline significant for Ki67 $p=0.0752$). Interestingly, expression of the

proliferation markers were lower in MGs of E2-exposed animals not bearing frank tumors (MG/T) than in both normal MG and in tumors (Figure 50A).

Hormone receptors showed interesting gene expression patterns during MG tumorigenesis (Figure 50B). While Pgr mRNA levels got significantly upregulated by E2 treatment, transcript levels of both, Esr1 and Esr2 were downregulated in tumors independent of the diet administered (not statistically significant). Expression levels in MG/T ranged between normal untreated MGs and E2-induced tumors. Results of the statistical analyses accounting for multiple tissues derived from individual animals are summarized in Table 9.

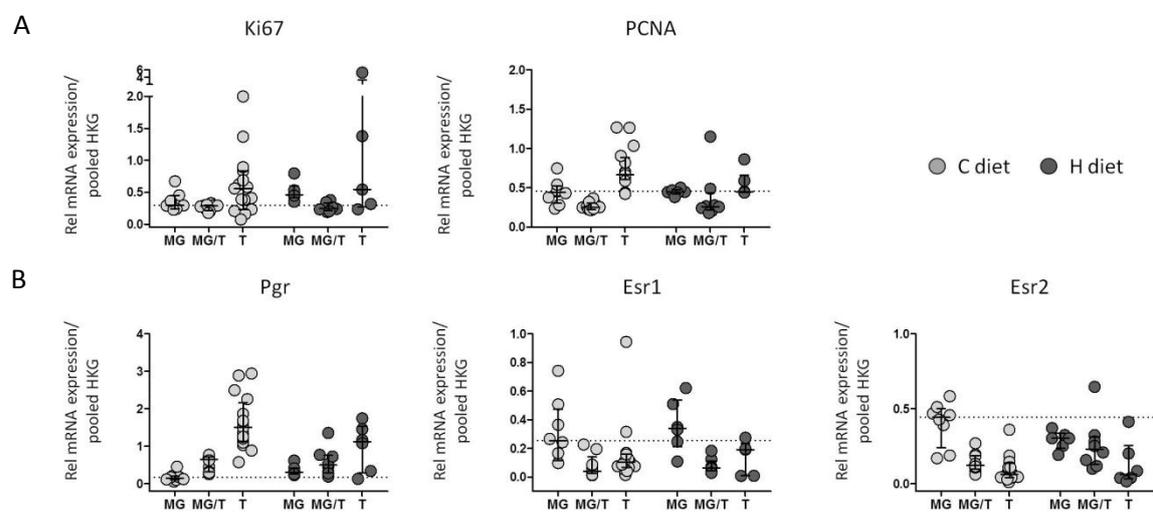


Figure 50: Gene expression of E2-responsive genes during MG tumorigenesis.

Relative gene expression of A: proliferation markers and B: hormone receptors in healthy mammary glands (MGs), MG without detectable tumors (MG/T) and tumors (T) at PND 285. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median expression levels in MGs in the C group. Vertical lines indicated interquartile range.

Table 9: Statistical analysis of gene expression during MG tumor development.

Source of Variation	Diet [p value]	Tissue global [p value]	MG+T vs. MG/T [p value]	Interaction [p value]
Ki67	ns	0.0752	ns	ns
PCNA	ns	<0.001	<0.01	ns
Pgr	ns	<0.001	<0.001	ns
Esr1	ns	ns	ns	ns
Esr2	ns	ns	ns	ns

Statistical significance of differences was calculated for all candidate genes using a linear mixed model. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. Models for diet and tissue effects as well for interaction were fitted. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction. Tissue effects were calculated across all tissue types (global: mammary glands (MGs) vs. mammary glands with no palpable tumors (MG/T) vs. tumors (T)) as well as MG and tumor compared to MG/T (MG+T vs. MG/T). ns: not statistically significant.

4.3.5.2 DNA methylome mining during tumor formation

In the carcinogenesis experiment, 24 samples were subjected for RRBS library preparation (MGs of C (n=5) and H (n=6) diet groups, tumor samples of C (n=6) and H (n=7) diet groups). We covered about 2 M CpGs (unique aligned motifs, Figure 51A). By sequencing libraries on two lanes (instead of one as in the previous experiment) we obtained 26-68 M high quality reads per sample with an alignment rate ranging from 66-88 % (except for 36 % for one tumor samples with was excluded from further analysis). In general, mean coverage was improved from 32x to 71x per site, when compared with libraries obtained from pre- and postpubertal ACI rats (see section 4.3.3). Also, average bisulfite conversion rate was increased from 98.5 to 99.0 % due to the optimized library preparation. Calculation of methylation data from all uniquely aligned motifs for MG and tumor samples of animals from PND 285 showed a mean global methylation of 53 %, in comparison to 64 % of the pre- and postpubertal ACI rats at PND 21 and 50 (Figure 51B).

A general summary of the obtained sequencing results of adult ACI rats can be found in Supplemental table 5.

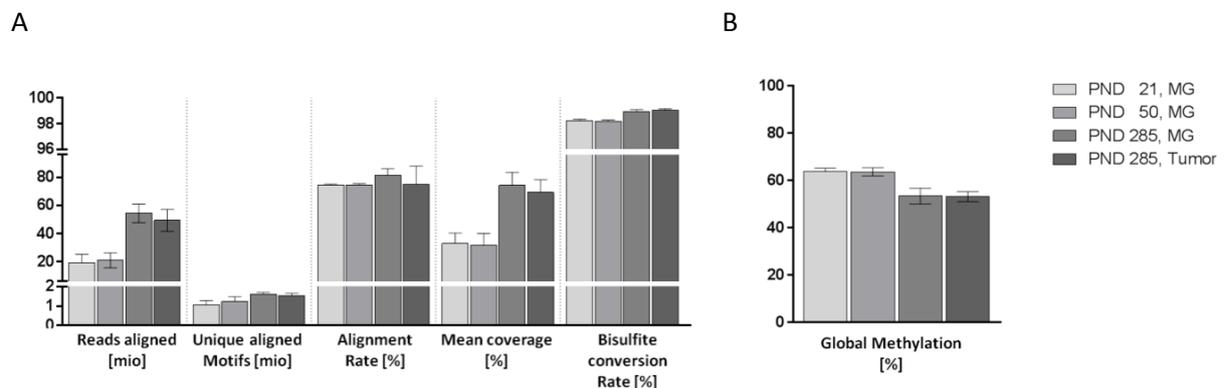


Figure 51: Summary of RRBS library runs from the puberty and the carcinogenesis study.

Comparison of RRBS library runs for general sequencing characteristics (A) and global methylation (B). PND 21 and 50: pre- and postpubertal ACI rats; PND 285: adult ACI rats, MG: Mammary gland

Genomic Distribution

For efficient detection of DMS between groups, only sites were considered with a minimum coverage of 20 reads in at least 20 out of the 24 samples within the groups, leading to 666,745 overlapping CpGs in all diet groups (Figure 52A).

Looking for significant methylation changes of at least 5 %, more than 40,000 sites were identified as differentially methylated as a result of the E2 treatment and the subsequent carcinogenic process in both, C and H diets, twice as many sites as due to IF intervention alone. During tumor development

(C vs. C+E2 and H vs. H+E2), we observed a genome-wide loss in methylation. Two thirds of the identified DMS accounted for hypomethylated sites with half of this sites mapping to repetitive genomic regions and might contribute to genome-wide instability [182] (Figure 52B). Gain in methylation was most prominent in intergenic and intronic regions (38 % and 20 %, respectively).

A

24.7 M CpGs				
Coverage (all)	2,27 M CpGs			
minCoverage >20 (20 out of 24)	666,745 CpGs			
Tissue	MG	MG	Tumor	Tumor
Diet / E2 intervention	C	H	C + E2	H + E2
	↓	↓	↓	↓
Main impact	Tumorigenesis	Tumorigenesis within IF diets	IF effects in MG	IF effects in tumors
comparison groups	C vs. C+E2	H vs. H + E2	C vs. H	C + E2 vs. H + E2
Total DMS	49,163	43,897	24,876	24,015
Hypomethylated	30,672	27,317	9,082	9,414

B

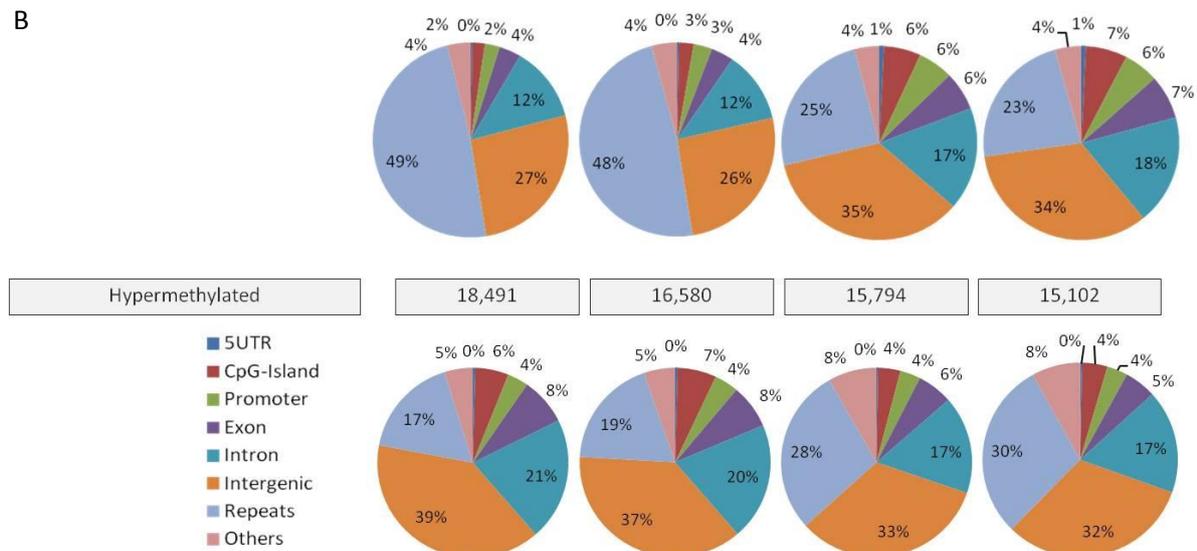


Figure 52: Flowchart and pie charts of genomic distribution of DMS in MGs and tumors from the carcinogenesis study.

A: Pipeline defining a series of cutoffs (minimum coverage of 20 reads in at least 20 out of the 24 samples within the four groups, $p < 0.05$ for group-wise comparison, methylation difference $> 5\%$) for efficient DMS detection. B: CpG sites were annotated to 8 categories (indicated by colors) according to the nearest transcription start site, and grouped in % relative to the total number of CpGs above each chart. Sites with loss in methylation are depicted as hypomethylated, sites with gain in methylation as hypermethylated DMS.

In contrast, effects of a lifelong IF exposure were less pronounced. In total ~24,000 CpGs were affected between the different diets when comparing normal MGs (C vs. H) or tumors (C+E2 vs.

H+E2). Here, H diet increased methylation of two thirds of all DMS but distribution of the sites was roughly equivalent. In general, less than 10 % of these selected sites were located in regulatory regions (promoters, CpG islands or 5' UTRs), with some overlap between groups as shown in a Venn diagram (Figure 53A).

In each group, more than 50 % of DMS were specifically affected by the particular treatment and were not shared with any other groups. The residual sites overlapped with at least one more treatment regime. The greatest overlap can be seen between the two tumor comparisons in C vs. C+E2 and H vs. H+E2 (blue vs. red) with 7,872 sites. Most of these sites mapped to repeat regions and most probably correspond to hypomethylated DMS (site not distinguished here).

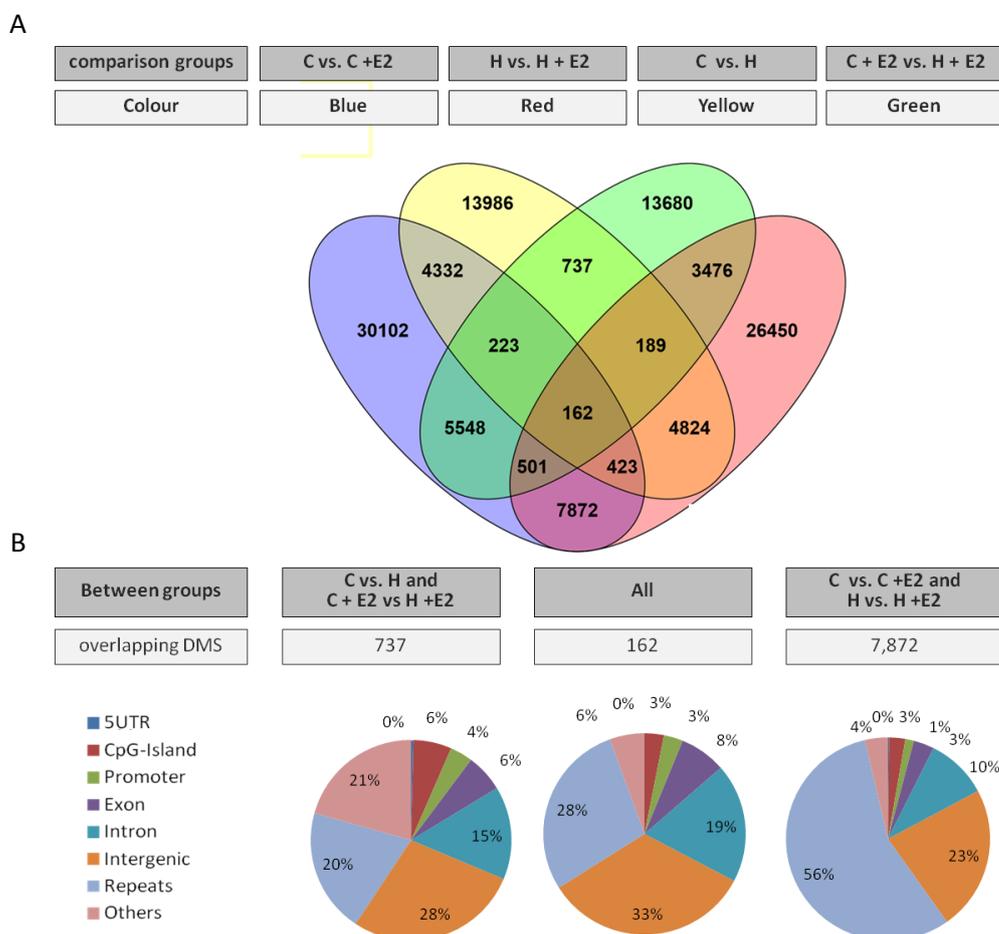


Figure 53: Venn diagram and genomic distribution of overlapping DMS from the carcinogenesis study.

A: Venn diagram of overlapping DMS between different group comparisons as indicated by colors. B: Genomic distribution of a subset of overlapping DMS. CpG sites were annotated to 8 categories (indicated by colors) according to the nearest transcription start site, and grouped in percent relative to the total number of CpGs above each chart.

A pure IF effect can be seen in the overlap between C vs. H and C+E2 vs. H+E2 (yellow vs. green). Many sites were shared also with the two other groups, but 737 sites were exclusively induced due

to the IF exposure. Interestingly, these sites were summarized in the category "others" where IF exposure especially affected methylation of ribosomal RNAs and short cytoplasmic RNAs (data not shown). A total of 162 DMS were shared by all groups. These sites might be important during normal MG development and are affected by IF and/or E2 intervention. The direction of DNA methylation change might vary between the groups and distribution of these sites throughout the genome did not show enrichment in a particular category.

The chemopreventive effects of lifelong IF exposure evolve in distinct biological processes

In order to understand the underlying mechanisms and the molecular events affected by a lifelong IF exposure during tumor development, we used the Core Analysis function included in the Ingenuity pathway analysis. DMS distinguishing tumors of the C vs. H diet were investigated to interpret IF effects on biological processes, signaling pathways and cell networks (Figure 54).

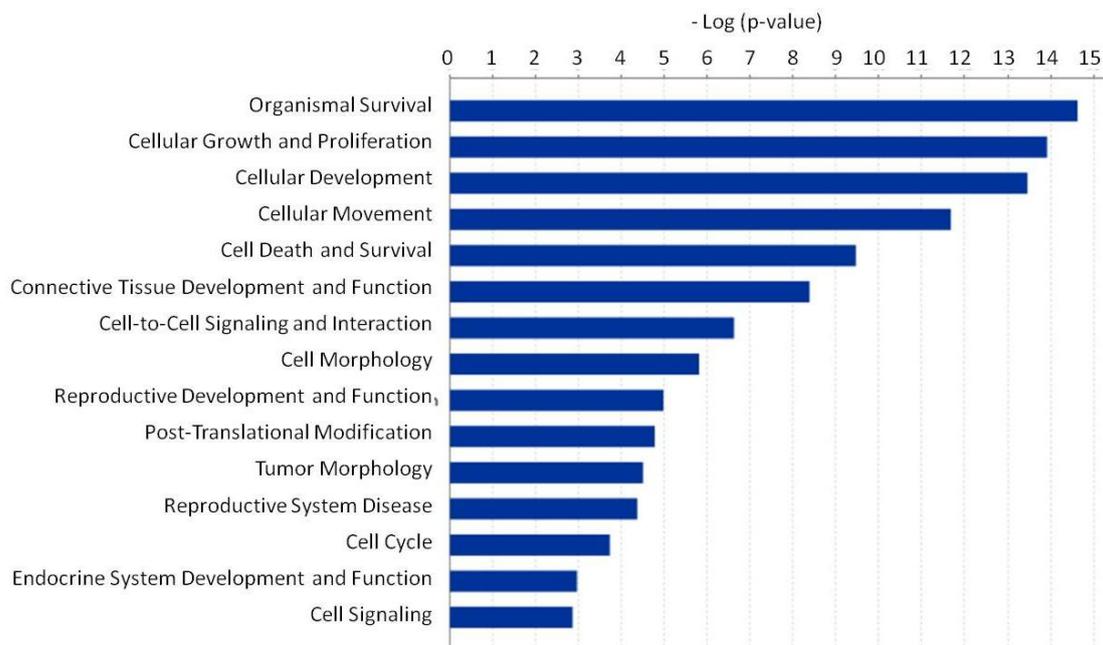


Figure 54: Functional gene enrichment based on DMS differentially methylated between tumors in the C vs. H diet groups.

Functional pathways with significantly enriched gene sets are ranked according to their p-values obtained with "Core Analysis" function using Ingenuity pathway analysis.

Therefore, the 24,015 IF-induced DMS that distinguish tumors of C and H diets (comparison C+E2 vs. H+E2, see Figure 52) were annotated to genomic identifiers and were defined as value parameters for the analysis: Hypomethylated genes were assumed to be downregulated and hypomethylated genes upregulated by translating methylation values into a -2 to 2 expression scale as Ingenuity pathway analysis is not able to work with methylation data directly. Both, hypo- and

hypermethylated DMS were filtered for close proximity to the next TSS (5 kb) to postulate a functional relationship between DMS and the annotated identifier. In total, 1565 affected genes were analyzed for significant enrichment in molecular, cellular and physiological functions as well as in diseases and disorders.

Using this approach, we identified that differentially methylated genes were highly enriched in functional pathways such as organismal survival, growth and proliferation, cellular development and movement, as well as cell death and survival, and many more with top hypermethylated genes, *i.e.*, RAX, SMAD7, MAFK, and Wnt5a as well as top hypomethylated genes, *i.e.*, PIK3R1, Trip11, and HOXD3 being involved in several of the observed processes.

4.3.5.3 Technical confirmation of quantitative methylation levels from RRBS

By sequencing of bisulfite converted DNA and subsequent bioinformatics analysis, RRBS provided quantitative methylation data in single nucleotide resolution. Nevertheless we aimed to confirm genome-wide data by using mass spectrometry-based EpiTYPER MassARRAY technology. In the previous study in pre- and postpubertal ACI rats, direct confirmation of RRBS-identified DMS was impeded by the fact that diet-induced differences were small and interindividual variation was high. By applying more stringent cutoff criteria for candidate site selection, we aimed to increase the chances to detect true DMS between groups.

Selection of candidate regions for RRBS confirmation in adult ACI rats

Candidate site selection from the RRBS data was based on a minimum coverage of 10x in every sample with a significant methylation difference of at least 5 %. A summary of our selection criteria is summarized in Figure 55.

Due to the library preparation and sequencing some animals did not meet the initial minimum coverage of 20x per CpG and values of these samples were set NA. In the next selection step just one NA per group was allowed and DMS for validation should be in close proximity to each other with less than 300 bp distance. After applying these settings, differential methylation was identified for 4,230 CpGs due to IF intervention and 12,816 CpGs as result of the E2 treatment (comparison 1 and 2, respectively).

Further E2-induced CpG sites were selected when we observed a concomitant preventive effects of the IF intervention (comparison 3). Here 12,816 E2-induced DMS were overlapped with 3,856 DMS found between the tumors of different dietary regimes (C+E2 vs. H+E2) resulting in a total of 271 CpGs that got differentially methylated during tumor development within the C diet but show an opposing methylation tendency in tumors of the H diet. These methylation levels might be more like

those in normal MGs and may underscore the preventive effect of the H diet in terms of DNA methylation changes during tumorigenesis.

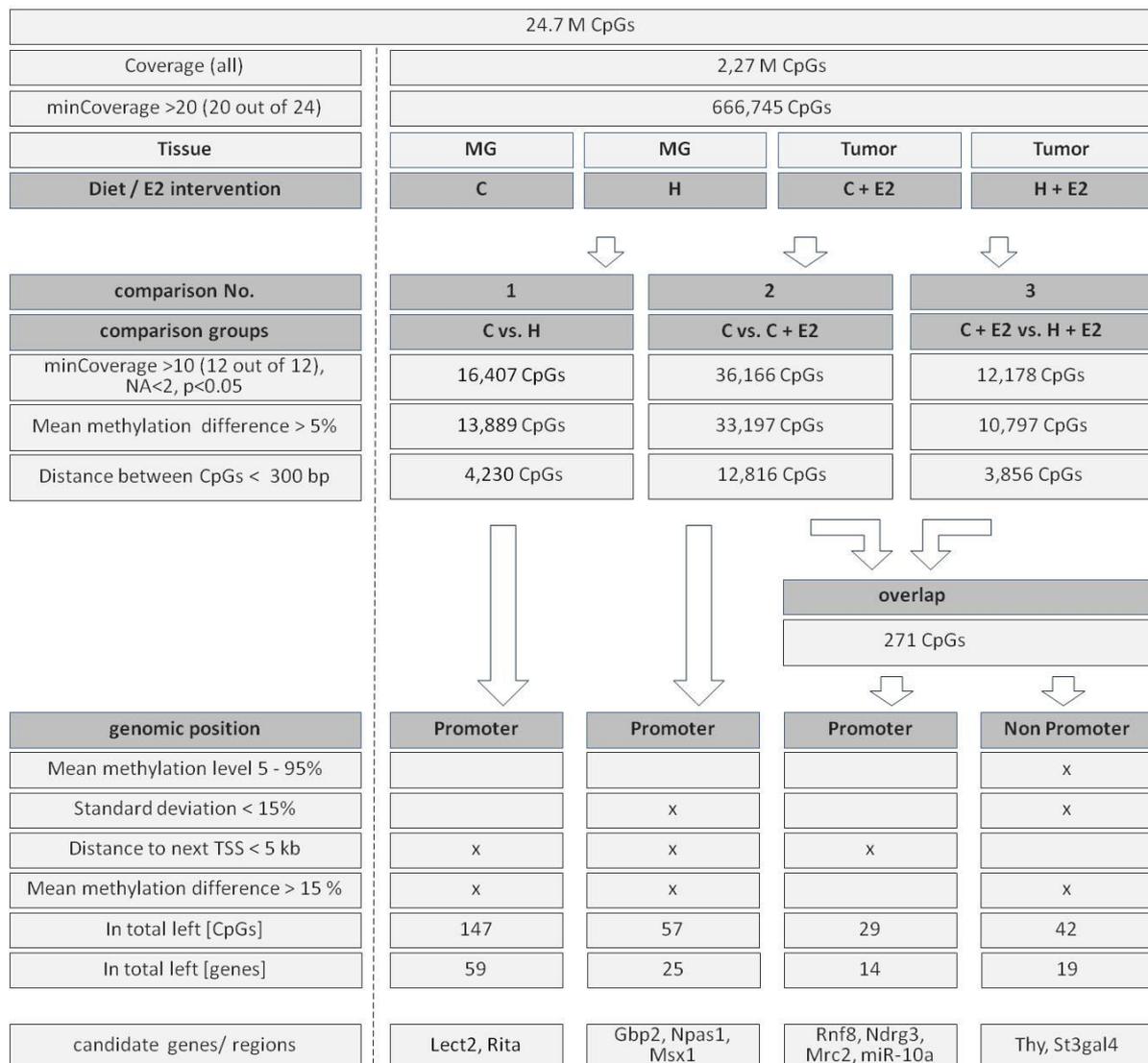


Figure 55: Flowchart for the selection of candidate regions for RRBS confirmation from the carcinogenesis study.

Pipeline defining a series of cutoff criteria (max. one NA value (<20x per CpG), min. coverage of at least 10x per CpG in all residual samples, p<0.05 for group-wise comparison, methylation difference >5 %) narrowing down the 24.7 M CpGs that are present in the rat genome for efficient DMS detection.

In order to enable a reasonable selection of candidate genes for conformation, the list of sites was further reduced by setting additional criteria. This approach was also used to investigate which setting might best identify those CpG sites with true methylation changes that could be confirmed by alternative methods. For the IF-induced effects in the MGs and E2-induced effects in the C diet group (comparison 1 and 2), regions of interest were narrowed down by the following setting: i) closer proximity to the next TSS with less than 5 kb distance and ii) methylation difference of >15 %. These

settings retained 147 sites in 59 genes affected by the IF intervention and 57 sites in 25 genes as a result of the E2 treatment. Latter sites were also filtered for standard deviation <15 % absolute methylation to exclude sites that were identified as DMS as result of individual outlier values.

In order to select regions that showed a preventive effect of the IF intervention (comparison 3), sites were either filtered for a less than 5 kb distance to the next TSS or i) for standard deviation <15 % absolute methylation, ii) methylation difference greater than 15 % and iii) mean methylation levels in the range of 5-95 %. The third setting was intended to particularly exclude sites which were identified as either unmethylated or 100 % methylated. These sites had been identified as false positives or negatives in the previous validation experiment in young ACI rats. In total 29 sites in 14 genes for promoter regions and 42 sites in 19 genes for non-promoter regions were retained.

Overall, eleven regions containing DMS mainly located in gene promoter regions or close to the TSS were selected to verify methylation with EpiTYPER MassARRAY in all experimental groups (Supplemental table 4).

Technical confirmation of RRBS data using MassARRAY for MG and tumor samples of ACI rats

Candidate regions selected for confirmation by MassARRAY were represented by at least 2 single CpGs or a CpG unit consisting of 2 CpGs. Methylation was measured in a zero-to-one scale.

For Comparison 1 (IF-induced methylation changes, C vs. H diet in MG samples), we did not set very stringent cutoff criteria, since evaluation of the RRBS data did not provide large numbers of H diet-induced DMS (see Figure 55). We selected DMS in the promoter regions of Lect2 (leukocyte cell-derived chemotaxin 2) and Rita (RBPJ interacting and tubulin associated 1). Lect2 has been linked to rheumatoid arthritis and amyloidosis [183]. However, in a recent work from Andres et al. it was identified amongst other as important predictor of breast cancer recurrence and mortality among smokers [184]. Overexpression of Rita upregulated p53 and reduced cyclinD1 and cyclinE levels, favoring apoptosis over cell proliferation in HepG2 cells and might function as a TSG in hepatocellular carcinoma [185, 186].

Selected candidate sites were often unmethylated (0 % methylation) such as for Lect2, or fully methylated (100 % methylation) such as for Rita in the RRBS dataset. These extreme methylation values could not be validated by MassARRAY and apparently represent technical and bioinformatics artifacts (see Figure 56A). So far, we do not fully understand, how generation of these artifacts can be prevented but bioinformatics optimization is ongoing.

For comparison 2 (carcinogenesis-induced methylation changes in the control group, C vs. C+E2), we took the information about artifacts into consideration by setting a threshold for standard deviation below 15 % methylation (see Figure 55). This filter removes highly variable CpGs with values ranging

from 0 to 100 % methylation within one group. We selected promoter regions for three out of 25 genes passing our thresholds, Npas1 (neuronal PAS domain protein 1), Msx1 (msh homeobox 1), and Gbp2 (guanylate binding protein 2, interferon-inducible).

The exact function of the PAS-family TF Npas1 is unclear, however it may play a role during late embryogenesis and postnatal development and has shown copy number variations in ER α -positive MCF-7 cells [187]. The homeodomain-containing TF Msx1 possesses repressive control on organogenesis and tissue interaction during embryonic development. It has been reported that promoter binding of Msx1 influences enrichment of H3K9me2 at repressed targets and redistributes the H3K27me3 mark, thereby affecting epigenetic regulation of its target genes [188]. The encoded protein of Gbp2 is a GTPase which has been associated with good prognosis in breast cancer [189].

RRBS data obtained for MGs and tumors of the C diet group showed significant hypomethylation for Gbp2 and considerable hypermethylation for Npas1 and Msx1 (Figure 56). Methylation differences between MGs and tumors of the C diet could be clearly validated. As an example, values for Gbp2 showed great concordance between the two quantitative methods with a slightly overall decrease in methylation for CpG 3 in the MassARRAY analysis. Overall, MassARRAY analyses revealed lower interindividual variation within the groups, indicating again that some of the more extreme methylation values detected by RRBS close to 0 or 100 % methylation might be artifacts.

In **comparison 3**, we were interested in carcinogenesis-induced methylation changes that could be prevented by intervention with IF (C vs. C+E2 vs. H+E2). Promoter (Figure 56C) and non-promoter regions (Figure 56D) were investigated comparing RRBS data for MGs vs. tumors of the C group overlapping MGs vs. tumors of H diet group. For comparison of promoter regions, we selected Rnf8 (ring finger protein 8, E3 ubiquitin protein ligase), Ndr3 (N-myc downstream regulated gene family member 3), Mrc2 (receptor, C type 2), and miR-10a.

RNF8 plays an important role in histone ubiquitination and chromatin remodeling during DNA damage response. Depletion of this protein causes cell growth inhibition and cell cycle arrest. Ndr3 is an androgen regulated aberrantly expressed gene in prostate cancer which might serve as a prognostic marker, since it is correlated with a shorter recurrence-free survival and shorter overall survival [190]. The encoded protein of Mrc2 is a central component in the collagen turnover process that directs basement membrane collagen as well as interstitial collagen to lysosomal degradation. Modulation of the tumor-surrounding extracellular matrix is a key determinant for cancer invasion, thus Mrc1 deregulation has been identified in various malignancies including breast cancer [191]. Deregulation of miR-10a has been reported in a number of cancers, including breast cancer and expression displayed both onco-miR as well as TSG roles by targeting the PTEN/AKT/ERK pathway. In

postmenopausal ER α -positive breast cancer patients, miR-10a upregulation was associated with longer relapse-free time following tamoxifen treatment and might rather possess TSG potential in he breast [192].

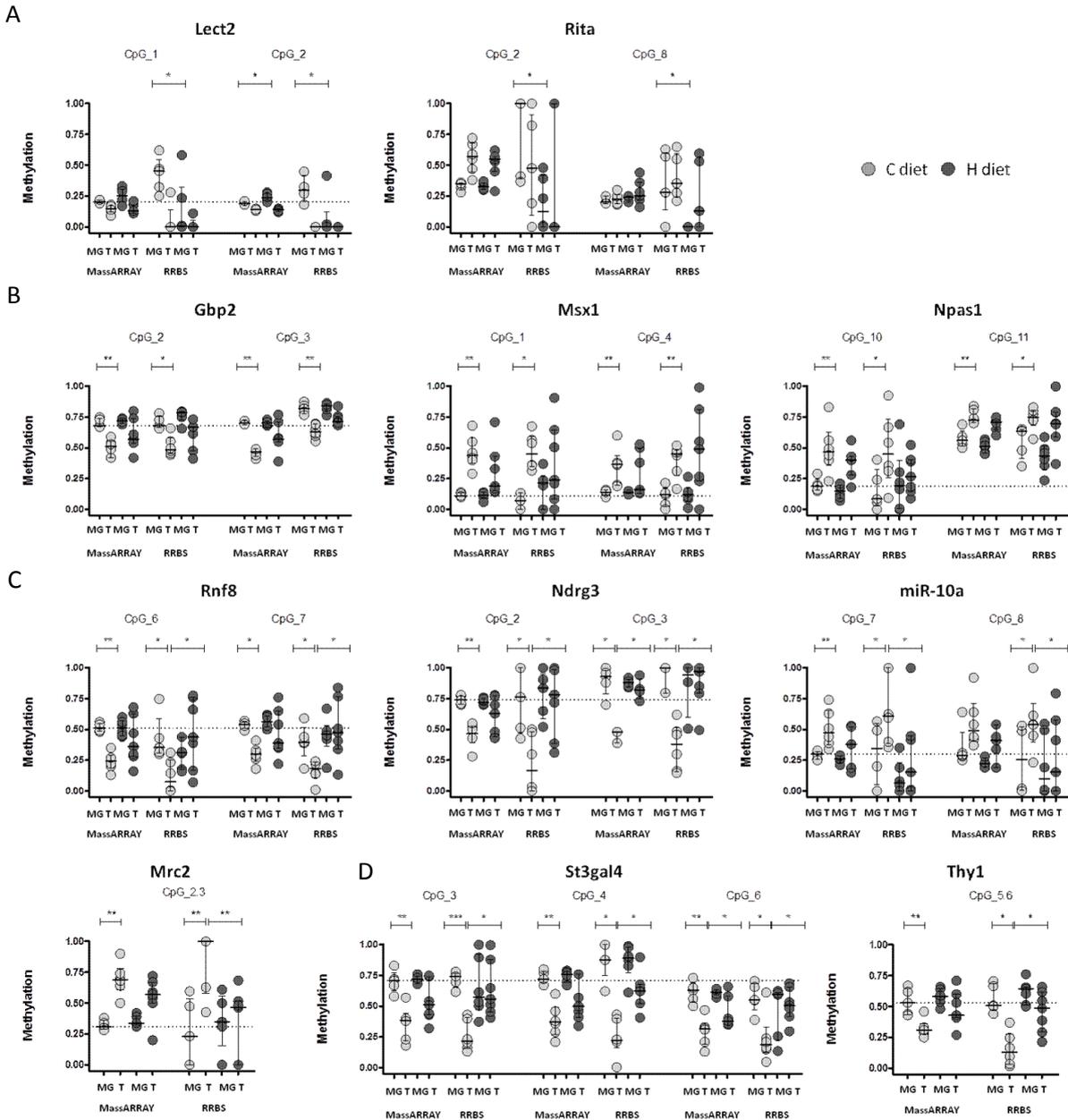


Figure 56: Technical confirmation of RRBS data by MassARRAY technology from the carcinogenesis study. DMS between A: MGs of the C vs. H diet (comparison 1), B: C vs. C+E2 diet (comparison 2) and C vs. C+E2 diet overlapping H vs. H+E2 diet (comparison 3) in C: promoter and D: non promoter regions. Each dot in the aligned dot plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed lines indicate median methylation values from the MassARRAY analysis of the first CpG of the control group. The vertical lines indicate inter-quartile range. MG: mammary gland, T: tumor

For the non-promoter comparison, we selected Thy (Thy-1 cell surface antigen), a surface glycoprotein expressed on multiple cell types, affecting cell-cell or cell-matrix interactions, as well as cellular adhesion and migration. Further, St3gal4 (ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (sialt4c)) passed our selection criteria. Sialylation plays a crucial role in cell adhesion and regulates the biological stability of glycoproteins by influencing the hydrophilic character and electronegative charge [193].

No thresholds were set for standard deviation and consequently intraindividual variation was high with up to 30 % absolute methylation difference for Rnf8, Ndr3, miR-10a, and Mrc2 between the RRBS and MassARRAY values. However, methylation differences between MGs and tumors of the C diet and the prevention of this alterations by the IF intervention could clearly be confirmed (Figure 56C). Comparable results were obtained for non-promoter regions of this comparison. Since these regions were selected using the most stringent filtering criteria to avoid error-prone data, with standard deviation below 15 %, mean methylation between 5 and 95 %, and methylation difference >15 % as outlined in Figure 30, methylation data for St3gal4 and Thy1 showed great concordance between the two quantitative methods (Figure 56D).

Overall, we were able to confirm the RRBS data by cutting down the interindividual variation within one group by reducing standard deviation or overall mean methylation levels. Since we did not apply any of these setting to verify the IF effect in MGs of C vs. H diet (comparison 1), we were not able to confirm these findings. Tumorigenesis induced greater methylation changes, and RRBS-identified DMS could be confirmed by the independent MassARRAY analysis. In general, variation was the greatest in tumors of the H diet independent of the method of analysis, reflecting the individual biological response to IF exposure during tumor development.

Correlation of MassARRAY and RRBS data of MGs and tumors of ACI rats

Correlation analysis was performed to determine the general consistency of the RRBS and MassARRAY data. For comparison, only the specific CpGs in the MassARRAY amplicon that covered the differential methylated DMS were considered. For each DMR, median methylation was calculated from the selected CpGs and RRBS methylation values were plotted against MassARRAY values in a 0 to 1 scale.

Notably, we could confirm that the 287 quantitative methylation values that were measured by both RRBS and MassARRAY significantly correlate (Spearman correlation coefficient ρ : 0.7868, 95% CI = 0.7368-0.8284 and $p < 0.0001$, Figure 57). This underscores that RRBS is a reliable tool to investigate

DNA methylation changes on a genome-wide scale, especially when specific cutoff criteria for selection are met (listed in Figure 55).

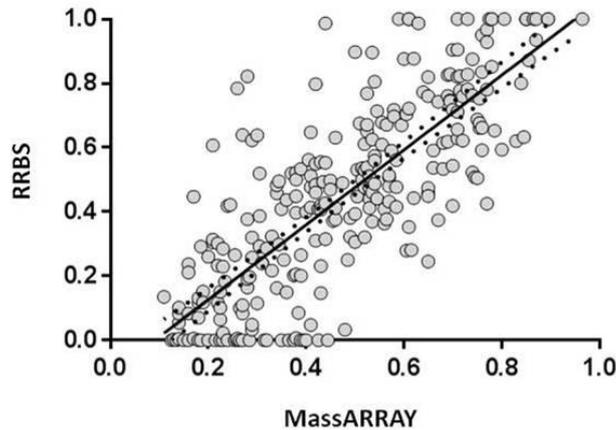


Figure 57: Correlation of RRBS and MassARRAY data generated from the carcinogenesis study.

Each dot in the correlation plots represents one median methylation values obtained from 24 animals in twelve candidate regions. In total, 287 quantitative methylation values measured by RRBS and MassARRAY show a significant correlation with a Spearman correlation coefficient ρ : 0.7868, 95% CI = 0.7368-0.8284 and $p < 0.0001$.

4.3.5.4 Quantitative analysis of candidate regions during MG tumor development

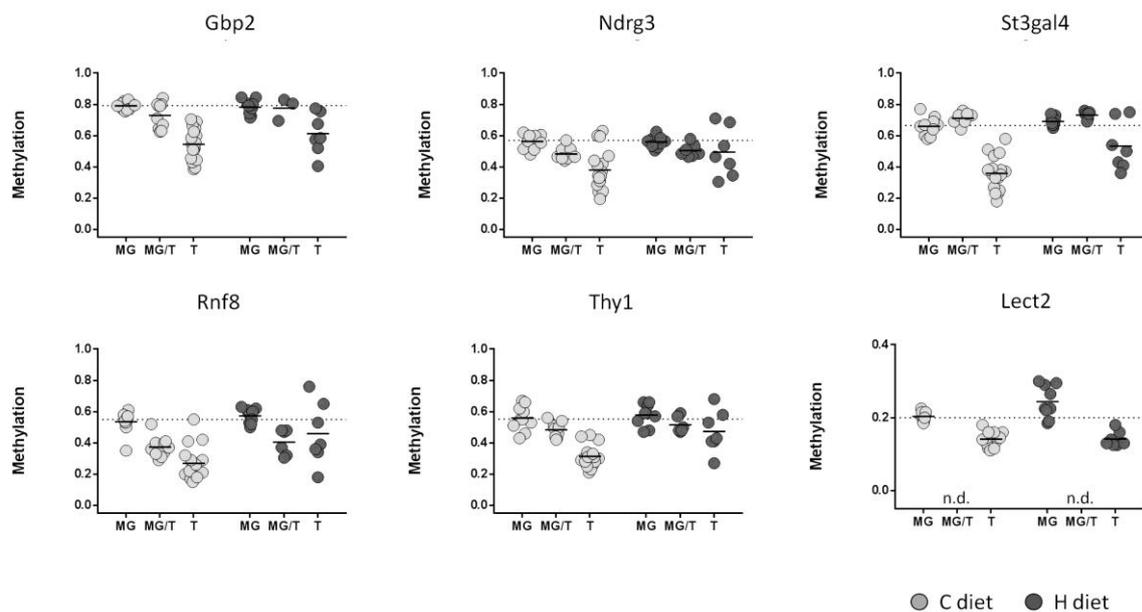
Next, we were interested to investigate methylation changes of validated candidate regions in the course of MG tumorigenesis, therefore we extended our quantitative analysis to all samples available from the carcinogenesis study. From C and H diet groups, we analyzed healthy MGs without E2 challenge (MG; $n=20$) and tumors from individual animals (T; $n=24$) as well as MGs from E2-exposed animals (MG/T; $n=17$) (Table 5).

In all regions investigated, we detected a significant impact of the tumorigenic process on methylation levels (Table 10, except for Rita). *Gbp2* and *Ndr3*, both used as markers for either a good prognosis in breast cancer or shorter overall survival in prostate cancer, became gradually hypomethylated by up to 25 %. Additionally, methylation levels of *Thy*, a glycoprotein involved in cellular adhesion and *St3gal4*, regulating glycoprotein stability decreased by up to 31 %. Also *Rnf8*, an E3 ubiquitin protein ligase involved in chromatin remodeling and *Lect2*, a predictor of breast cancer recurrence became hypomethylated during mammary carcinogenesis and methylation levels gradually decreased by up to 31 % (Figure 58).

Table 10: Statistical analysis of DNA methylation changes of candidate regions during MG tumor development.

Source of Variation	Diet [p value]	Tissue global [p value]	MG+T vs. MG/T [p value]	Interaction [p value]
Gbp	ns	<0.0001	ns	ns
Ndr3	ns	<0.01	ns	ns
St3gal4	ns	<0.0001	ns	ns
Rnf8	ns	<0.0001	<0.01	ns
Thy1	ns	<0.0001	ns	ns
Lect2	0.0972	<0.0001	ns	<0.05
miR-10a	ns	<0.0001	0.0904	ns
Npas1	ns	<0.0001	0.0864	ns
Mrc2	ns	<0.0001	ns	ns
Msx1	ns	<0.0001	ns	ns
Rita	ns	ns	ns	ns

Statistical significance of differences was calculated for all candidate genes using a linear mixed model. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. Models for diet and tissue effects as well for interaction were fitted. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction. Tissue effects were calculated across all tissue types (global: mammary glands (MG) vs. mammary glands with no palpable tumors (MG/T) vs. tumors (T)) as well as MG and tumor compared to MG/T (MG+T vs. MG/T).

**Figure 58: DNA hypomethylation of candidate regions during MG tumor development.**

Median methylation in healthy mammary glands (MGs), mammary glands with no palpable tumors (MG/T) and tumors (T) from PND 285 analyzed by MassARRAY. Median methylation of MGs of the C diet is indicated by a horizontal dashed line. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, IF diet). For statistical analysis see Table 10.

In contrast, promoters of miR-10a, a TSG in breast cancer, the embryonic TF Npas1 and Msx1 as well as Mrc2, a gene involved in collagen turnover and often deregulated in breast cancer gained methylation, and methylation levels significantly increased by about 15-30 % in the tumor samples of the C group (Figure 59).

Interestingly, lifelong IF intervention induced Lect2 methylation in MG samples of the H group (borderline significant $p=0.0972$) and prevented the carcinogenesis-mediated loss or gain in methylation by 7-14 % for Rnf8, Thy, miR-10a, Npas1, and St3gal4, as well as Lect2. However, according to the statistical analysis, most of these preventive effects did not reach statistical significance (except for Lect2 $p<0.05$).

Particularly interesting are methylation levels of the MG/T samples. Unfortunately, no histopathological data is available to estimate the degree to transformation of these samples. By analyzing DNA methylation, we found that MG/T samples ranged between normal (untreated) MGs and E2-induced tumors, but could not statistically be distinguished from one or the other (except for Rnf8 $p<0.01$, borderline significant for miR-10a and Npas1 with $p=0.0904$ and 0.0864 , respectively). However, this data indicates that adjacent MGs without palpable tumors from rats treated with E2 cannot be considered as normal MGs as they already show tumor-characteristic DNA methylation patterns.

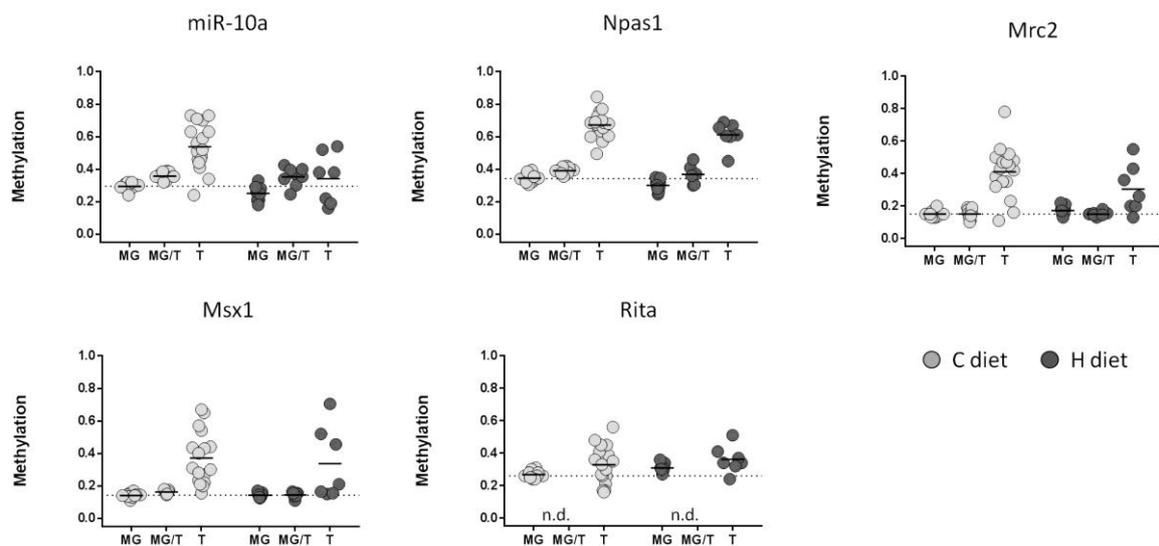


Figure 59: DNA hypermethylation of candidate regions during MG tumor development.

Median methylation in healthy mammary glands (MGs), mammary glands with no palpable tumors (MG/T) and tumors (T) from PND 285 analyzed by MassARRAY. Median methylation of MG of the C diet is indicated by a horizontal dashed line. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, IF diet). For statistical analysis see Table 10.

For Msx1, Mrc2 and St3gal4 we did not detect this gradual development of aberrant methylation. MG/T samples rather exhibited methylation patterns similar to healthy MGs. Eventually, epigenetic alterations of these regions might be a terminal event during tumor development and are therefore not yet visible in MG/T samples.

In addition to these regions selected from the RRBS dataset, we were interested whether regions selected from the RRBS investigation of pre- and postpubertal ACI rats (see section 4.3.3.2) would undergo additional methylation changes during E2-induced carcinogenesis (Figure 60, Table 11).

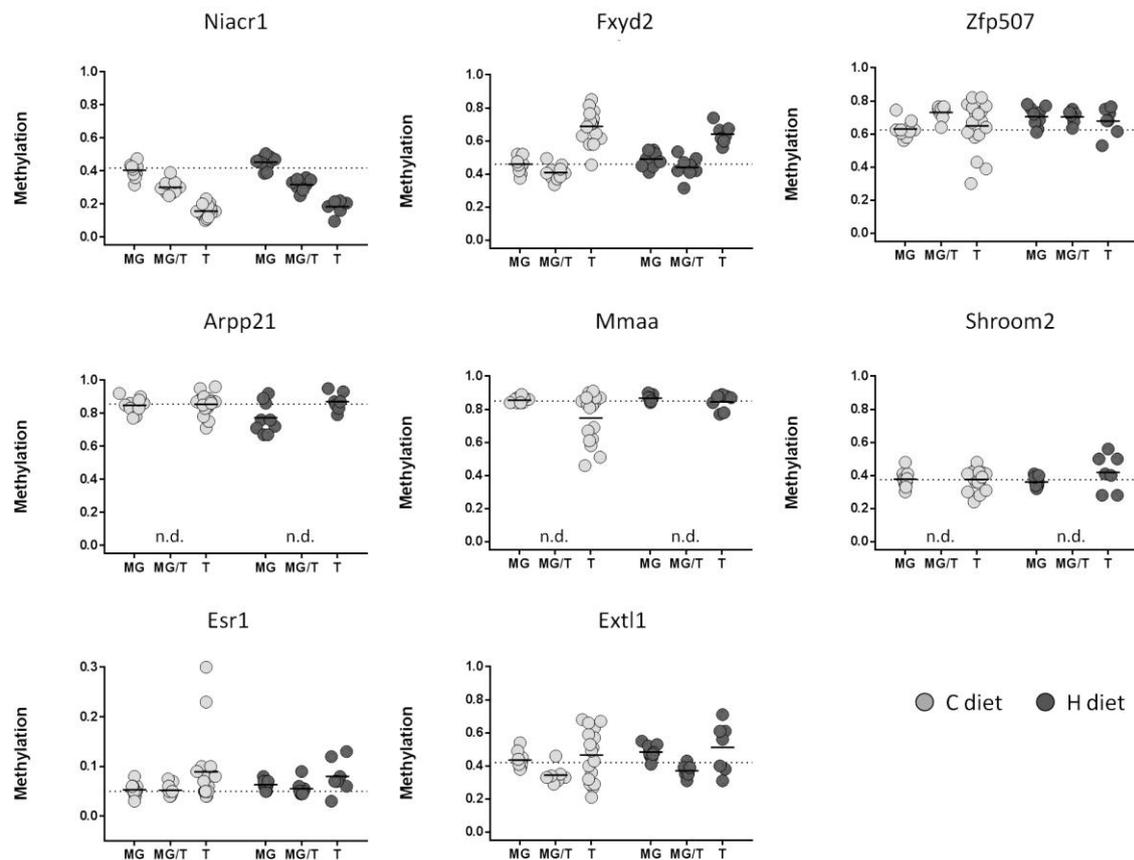


Figure 60: DNA methylation of candidate regions from the puberty study during MG tumor development.

Median methylation in healthy mammary glands (MG), mammary glands with no palpable tumors (MG/T) and tumors (T) from PND 285 analyzed by MassARRAY. Median methylation of MG of the C diet is indicated by a horizontal solid line. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, IF diet). For statistical analysis see Table 11.

Out of eight analyzed regions, methylation levels of *Niacr1* and *Fxyd2*, two genes involved in adipose tissue maintenance and energy homeostasis, as well as *Extl1*, an E2-inducible gene further changed in the course of carcinogenesis. Promoter methylation of *Niacr1* significantly decreased by about 30 % in both dietary groups. Promoter methylation of *Extl1* initially decreased by 5-10 % in MGs of E2-treated animals (MG/T) (borderline significant $p=0.0989$) before the region underwent a tumorigenesis-induced gain in methylation to control levels. Similar results have been obtained for *Esr1* and *Fxyd2*, methylation levels in MG/T did not change from control but promoter methylation increased by 5 % and 30 %, respectively during tumor development. For *Arpp21* and *Mmaa*, we focused on MG and T, only. Statistical analysis identified a borderline significant diet-induced

hypomethylation in the MGs of Arpp21 promoter ($p=0.0751$) when compared to control group. Notably, median methylation of Mmaa was especially homogeneous in the H group.

Zfp507 and Shroom2 showed less pronounced alterations and did not show a significant impact of either E2 treatment or lifelong IF intervention on promoter methylation levels at PND 285.

Table 11: Statistical analysis of DNA methylation of candidate regions during MG tumor development.

Source of Variation	Diet [p value]	Tissue global [p value]	MG+T vs. MG/T [p value]	Interaction [p value]
Niacr1	ns	<0.0001	<0.0001	ns
Fxyd2	ns	<0.0001	ns	ns
Zfp507	ns	ns	ns	ns
Arpp21	ns	0.0751	ns	ns
Mmaa	ns	0.0705	ns	ns
Shroom2	ns	ns	ns	ns
Esr1	ns	0.07	ns	ns
Extl1	ns	<0.05	0.0989	ns

Statistical significance of differences was calculated for all candidate regions using a linear mixed model. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. Models for diet and tissue effects as well for interaction were fitted. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction. Tissue effects were calculated across all tissue types (global: mammary glands (MG) vs. mammary glands with no palpable tumors (MG/T) vs. tumors (T)) as well as MGs and tumors compared to MG/T (MG+T vs. MG/T).

4.3.6 Gene expression analyses (performed by Karin Klimo)

4.3.6.1 Candidate gene expression and correlation with DNA methylation changes

Since we had selected candidate regions in the promoter regions of genes, we were interested whether changes in DNA methylation might be correlated with changes in gene expression. We aimed to analyze candidate genes obtained from the RRBS runs of pre- and postpubertal (Niacr1, Fxyd2, Arpp21, Shroom2, Zfp507, Mmaa), adult ACI rats (Ndr3, Lect2, Mrc2, Npas1, Msx1, Rita, Gbp2, Thy1, Rnf8, St3gal4, mirR-10a), as well as two candidates from the Wistar rat study (Esr1, Extl1).

No expression data could be obtained for Mmaa and Shroom2 due to unsuccessful primer design. Also, Lect2, Npas1, Msx1, Rita, Gbp2, and miR-10a mRNA levels could not be determined. These genes were previously reported as being not expressed in adult ACI rats [194].

Gene expression levels were therefore measured for Ndr3, Mrc2, Thy, Rnf8, St3gl4, Niacr1, Fxyd2, Arpp21, Zfp507, Esr1, as well as Extl1, and Spearman correlation coefficient rho was calculated for correlation of methylation and gene expression.

For six of these eleven genes we detected a significant impact of the tumorigenic process on mRNA expression (Table 12). Arpp21, Fxyd2, Rnf8, and Zfp507 mRNA transcript levels decreased during mammary carcinogenesis (Figure 61). Gene expression of Mrc2 initially decreased in MG/T before it underwent a tumorigenesis-induced upregulation close to control levels (borderline significant $p=0.077$).

Table 12: Statistical analysis of gene expression changes during MG tumor development.

Source of Variation	Diet [p value]	Tissue global [p value]	MG+T vs. MG/T [p value]	Interaction [p value]
Arpp21	<0.05	<0.001	<0.001	ns
Fxyd2	0.0898	<0.01	<0.01	ns
Rnf8	0.0590	<0.001	ns	ns
Zfp507	ns	<0.05	ns	ns
Esr1	ns	ns	ns	ns
Mrc2	ns	0.0773	ns	ns
Thy1	ns	<0.05	ns	<0.05
Ndr3	0.0505	ns	ns	ns
St3gal4	ns	<0.001	<0.01	ns
Niacr1	ns	ns	ns	ns
Extl1	ns	ns	ns	ns

Statistical significance of differences was calculated for all candidate genes using a linear mixed model. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. Models for diet and tissue effects as well for interaction were fitted. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction. Tissue effects were calculated across all tissue types (global: mammary glands (MG) vs. MGs with no palpable tumors (MG/T) vs. tumors (T)), as well as MGs and tumors compared to MG/T (MG+T vs. MG/T).

Thy expression was significantly induced by IF intervention during the carcinogenic process, with mRNA levels being significantly higher in the tumor samples of the H group than of the C group (Figure 62). Interestingly, St3gal4 mRNA expression was highest in MG/T without palpable tumors independent of the dietary regime.

Lifelong IF intervention led to increased transcript levels of Arpp21, Fxyd2, Rnf8 (Figure 61), and Ndr3 (Figure 62), borderline significant for Fxyd2, Rnf8, and Ndr3 $p=0.0505-0.0898$ in healthy MG of the H group. Expression levels in MG/T ranged between normal untreated MGs and E2-induced tumors but were distinct from MGs and tumors for Arpp21, Fxyd2 (Figure 61), and St3gal4 mRNA.

In six out of eleven candidate genes, methylation levels correlated with expression levels, with Spearman correlation coefficients rho in the range of 0.09 to 0.48 for positive correlation (St3gal4, Esr1 and Niacr1) and -0.36 to -0.37 for negative correlation (Ndr3, Arpp21, Fxyd2) with higher methylation leading to reduced gene expression (Figure 63). Correlation was weak for Esr1, since no significant methylation changes were identified during carcinogenesis. Correlations for Niacr1 and Fxyd2 were border line significant ($p=0.0809$ and $p=0.0710$, respectively).

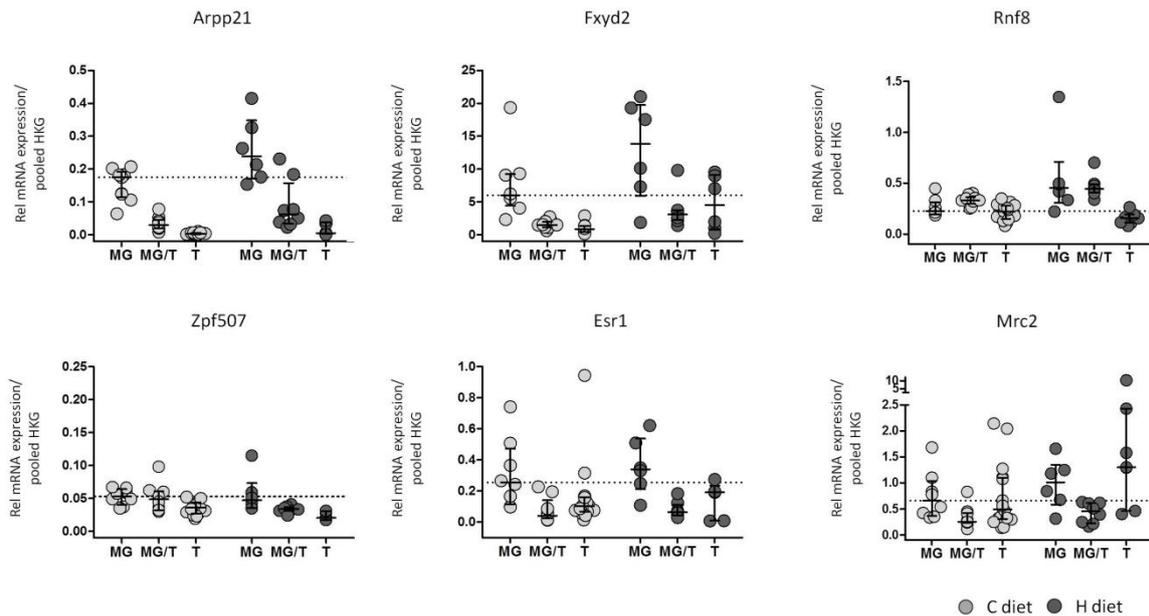


Figure 61: Expression of candidate genes decreasing during MG tumor development.

Relative mRNA expression to pooled housekeeping genes (HKG: Hprt, TBP and β -actin) in healthy mammary glands (MGs), MGs with no palpable tumors (MG/T) and tumors (T) from PND 285 of candidate regions obtained by RRBS. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median expression level of the MGs of the C diet. Vertical lines indicated interquartile range. For statistical analysis see Table 12.

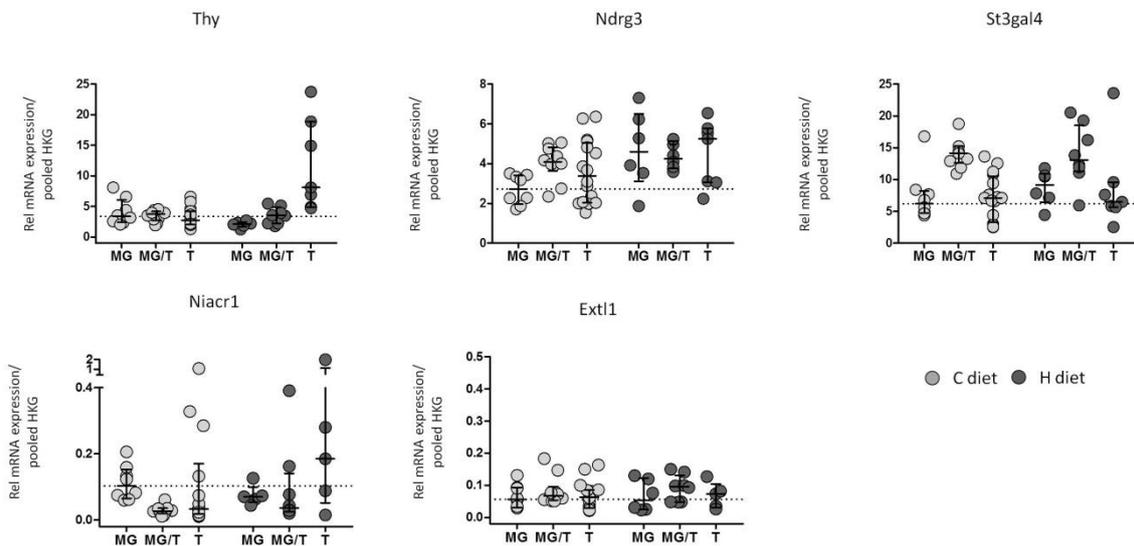


Figure 62: Gene expression changes during MG tumor development.

Relative mRNA expression to pooled housekeeping genes (HKG: Hprt, TBP and β -actin) in healthy mammary glands (MGs), MGs with no palpable tumors (MG/T) and tumors (T) from PND 285 of candidate regions obtained by RRBS. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median expression level of the MGs of the C diet. Vertical lines indicated interquartile range. For statistical analysis see Table 12.

Interestingly, all DMRs for which methylation levels correlated with expression levels were annotated to promoter regions. Indeed, DNA methylation might be responsible for gene silencing in these regions by either recruiting MBD proteins and HDACs, leading to subsequent chromatin compaction, or by inhibiting methylation-sensitive TF binding to the promoter regions. Also positive correlations with higher methylation leading to enhanced gene expression have been frequently reported. It has been proposed that the H3K36me3 mark which is associated with transcriptional elongation actively recruits DNMTs thus consequently increase methylation status, although its effect has been mainly limited to the intragenic regions [195].

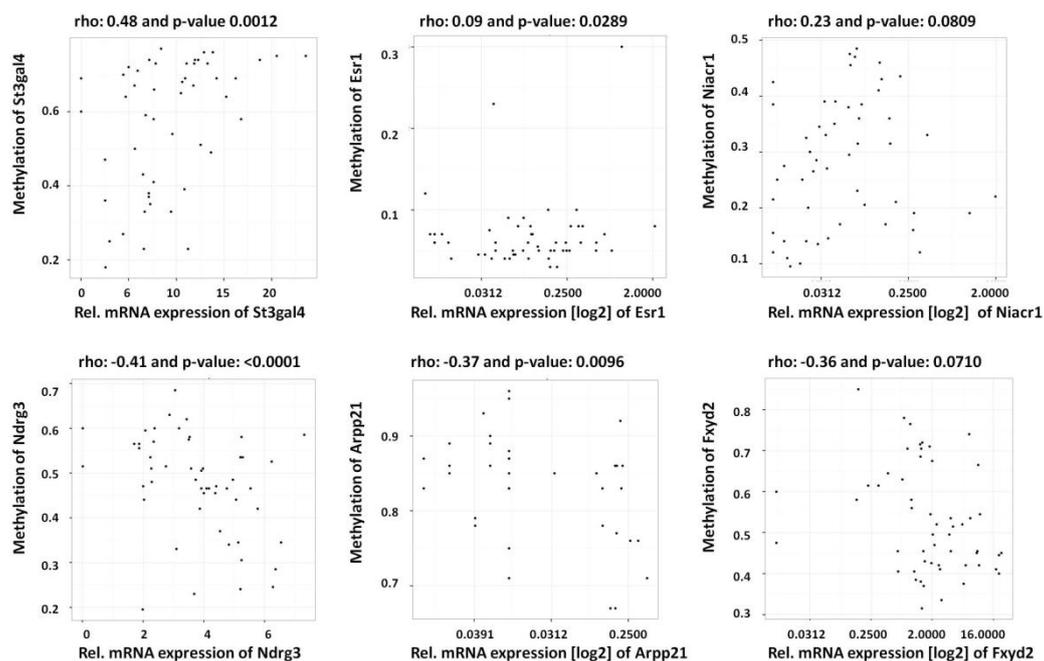


Figure 63: Correlation of DNA methylation levels and gene expression from the carcinogenesis study. Each dot represents one sample. Spearman correlation coefficient ρ was computed for the comparison of promoter methylation and gene expression.

4.3.6.2 Gene expression-based cell type characterization using reference markers

Both, methylation as well as gene expression analyses were performed with homogenates of whole organs/tumors representing a mix of cell types. Therefore, we characterized the cellular composition of healthy MGs, MGs without palpable tumors and tumors to be able to better interpret differential expression patterns, observed in normal MGs and MG tumorigenesis. We determined relative transcript levels for selected marker genes for epithelial cells (*Krt8* and *Cdh1*) and myoepithelial cells (*Krt5*, *SMA* and *Tp63*).

Due to the breast carcinogenic process, markers of epithelial cell differentiation, *i.e.*, Krt8 and Cdh1 got massively upregulated while myoepithelial markers, *i.e.*, Krt5, SMA and TP63 were down-regulated (nonsignificant for Krt5), indicating the enormous clonal expansion of luminal cells in mammary tumors (Figure 64, Table 13).

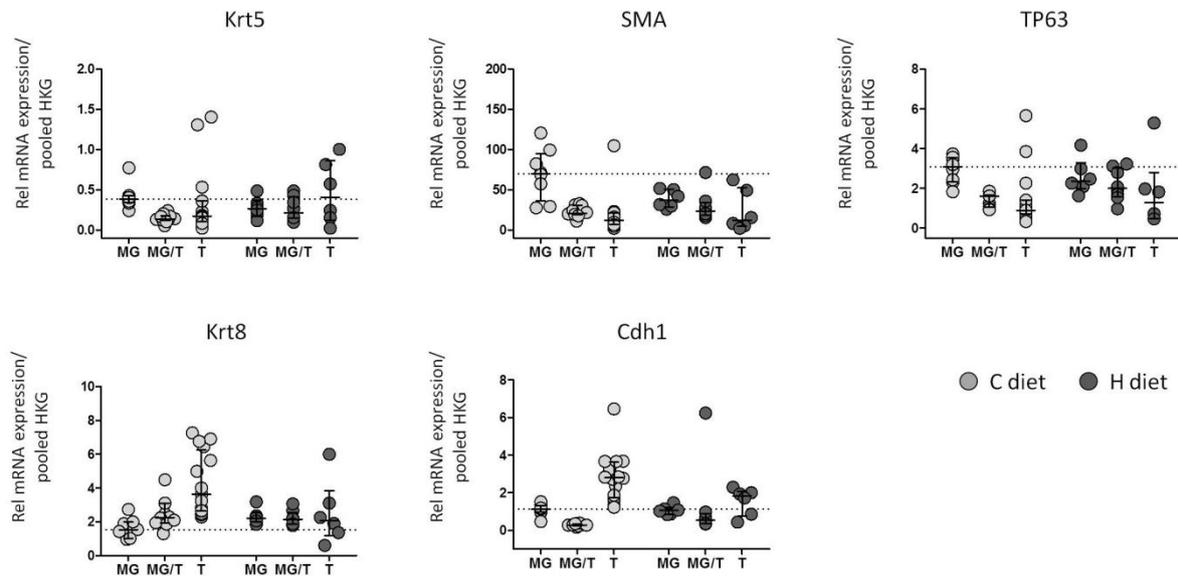


Figure 64: Gene expression of cell type markers during MG tumor development.

Relative mRNA expression to pooled housekeeping genes (HKG: Hprt, TBP and β -actin) in healthy mammary glands (MGs), MGs with no palpable tumors (MG/T) and tumors (T) from PND 285 of candidate regions obtained by RRBS. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median expression level of the MGs of the C diet. Vertical lines indicated interquartile range. For statistical comparison see Table 13.

Table 13: Statistical analysis of gene expression of cell type markers during MG tumor development.

Source of Variation	Diet [p value]	Tissue global [p value]	MG+T vs. MG/T [p value]	Interaction [p value]
Krt5	ns	ns	ns	ns
SMA	ns	<0.01	<0.05	ns
TP63	ns	<0.05	0.0538	ns
Krt8	ns	<0.01	ns	ns
Cdh1	ns	<0.01	ns	ns

Statistical significance of differences was calculated for all candidate genes using a linear mixed model. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. Models for diet and tissue effects as well for interaction were fitted. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction. Tissue effects were calculated across all tissue types (global: mammary glands (MGs) vs. MGs with no palpable tumors (MG/T) vs. tumors (T)), as well as MGs and tumors compared to MG/T (MG+T vs. MG/T).

Again, expression levels of mammary gland without palpable tumors range between normal untreated MG and E2-induced tumors but were distinct from MG and tumor for SMA and TP63

(borderline significant for TP63 $p=0.0538$). Gene expression of *Cdh1* initially decreased in MG/T before it underwent a tumorigenesis-induced upregulation above control levels.

No effect of the lifelong IF intervention was observed for the transcript levels of any cell marker.

4.3.6.3 IF affect gene expression of key epigenetic enzymes

In order to identify the underlying mechanism of IF and age-induced DNA methylation and gene expression changes during MG development, we measured relative mRNA expression of selected epigenetic modifying enzymes involved in writing and erasing DNA methylation (*e.g.*, DNMT1, 3a, 3b and Tet1) (Figure 65 and Table 14).

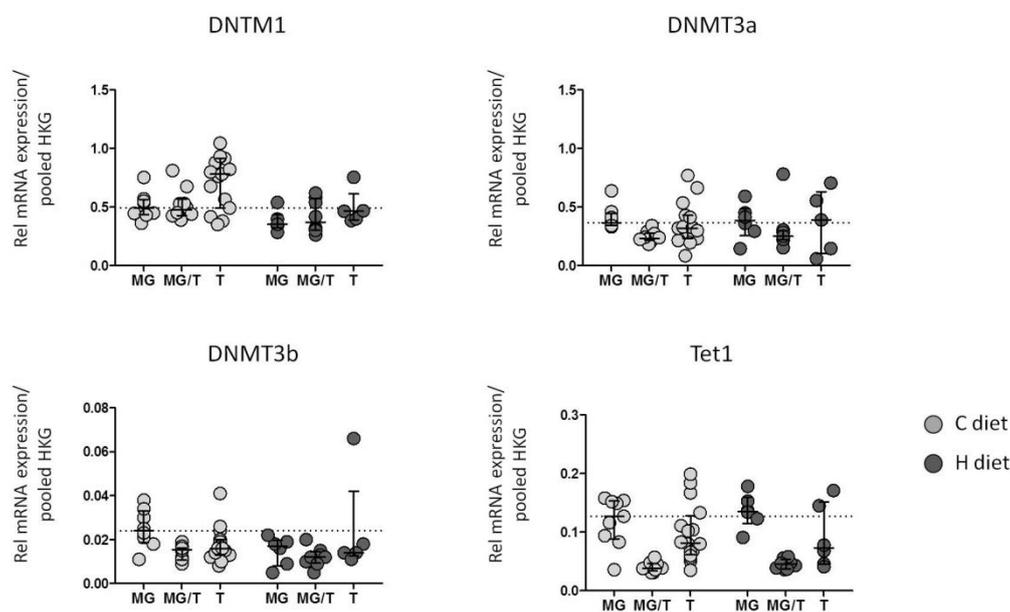


Figure 65: Gene expression of epigenetic key enzymes during MG tumor development.

Relative mRNA expression to pooled housekeeping genes (HKG: *Hprt*, *TBP* and β -actin) of epigenetic key enzymes in healthy mammary glands (MGs), mammary glands with no palpable tumors (MG/T) and tumors (T) from PND 285. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, H diet). The horizontal dashed line indicates median expression level of the MG of the C diet. Vertical lines indicated interquartile range. For statistical comparison see Table 14.

Similarly to the developmental effects, relative mRNA expression of DNA methylation modifying enzymes showed a diverse pattern. DNMT3a and DNMT3b mRNA levels were not significantly affected during tumor formation. Transcripts levels of DNMT1 and Tet1 showed opposing alterations in mRNA expression. While DNMT1 mRNA levels were upregulated in tumors, especially in the C diet group, Tet1 expression was decreased, with most pronounced reductions in MGs without palpable tumors (MG/T) independent of the dietary group. Interestingly, lifelong IF exposure significantly reduced mRNA levels of DNMT1 in healthy MGs of the H group as well as in MG/T and T and prevented the E2-induced upregulation of DNMT1 mRNA during tumor formation.

Table 14: Statistical analysis of gene expression of epigenetic key enzymes during MG tumor development.

Source of Variation	Diet [p value]	Tissue global [p value]	MG+T vs. MG/T [p value]	Interaction [p value]
DNMT1	<0.05	<0.05	ns	ns
DNMT3a	ns	ns	ns	ns
DNMT3b	ns	ns	ns	ns
Tet1	ns	<0.001	<0.001	ns

Statistical significance of differences was calculated for all candidate genes using a linear mixed model. A random rat effect was included to account for multiple tumor/tissue samples from the same animal. Models for diet and tissue effects as well for interaction were fitted. P-values were adjusted for multiple testing across genes using Benjamini-Hochberg correction. Tissue effects were calculated across all tissue types (global: mammary glands (MGs) vs. mammary glands with no palpable tumors (MG/T) vs. tumors (T)), as well as MGs and tumors compared to MG/T (MG+T vs. MG/T).

4.3.7 DNA methylation changes of enriched MG epithelial cells obtained by LCM

Since MGs are composed of multiple cell types, *i.e.*, ductal epithelial cells, adipocytes, stromal cells, immunological cells, and others, differences in methylation might reflect differences in cell composition between samples rather than alterations in methylation levels.

We assumed that methylation differences within mammary epithelial cells were “diluted” by the contribution of all cell types to an average methylation level as we used whole MG tissue for our quantitative analyses. In order to enrich for mammary epithelial cells from the cell mix present in a bulk MG, non-contact laser capture microdissection (LCM) was performed from MGs (n=6) of the C and H diets as well as DCIS (n=3) of the control group (Table 5). A two-step process of cutting the specimen and catapulting into adhesive cap tubes, without any contact and against gravity enabled an enrichment of the area of choice (Figure 66). For healthy MGs between 3-5 mm² and roughly 1 mm² from DCIS samples were used for DNA isolation.

In order to gain a clear insight into the DNA methylation levels of enriched MG epithelial cells, we analyzed candidate regions identified by RRBS analyses of pre- and postpubertal ACI rats as well as during ACI carcinogenesis by MassARRAY.

Due to limited amounts of DNA (mostly less than 30 ng), no PCR product was obtained for Niacr1, Arpp21, Fxyd2, and Mmaa. Interestingly, Extl1, Ndr3, Thy, and in part Esr1, and Zfp507 showed relative homogenous methylation levels within one dietary group, reflecting the observed median methylation when analyzing the whole MGs (Figure 67). With an exception of Thy and Zfp507, methylation levels determined for the LCM-samples was increased when compared to the non-dissected counterparts (with high variation for Zfp507 in the H diet) suggesting that the epithelial-surrounded cell types indeed are able to influence DNA methylation analysis.

In contrast to our expectations, interindividual DNA methylation was often greater in the LCM samples when compared to samples obtained from the MG bulk of cells. Especially miR-10a, Mrc2, Rnf8, St3gal4, and Npas1 showed massive divergence within one diet group which did not enable a clear insight into epithelial DNA methylation changes unbiased by other cell subtypes.

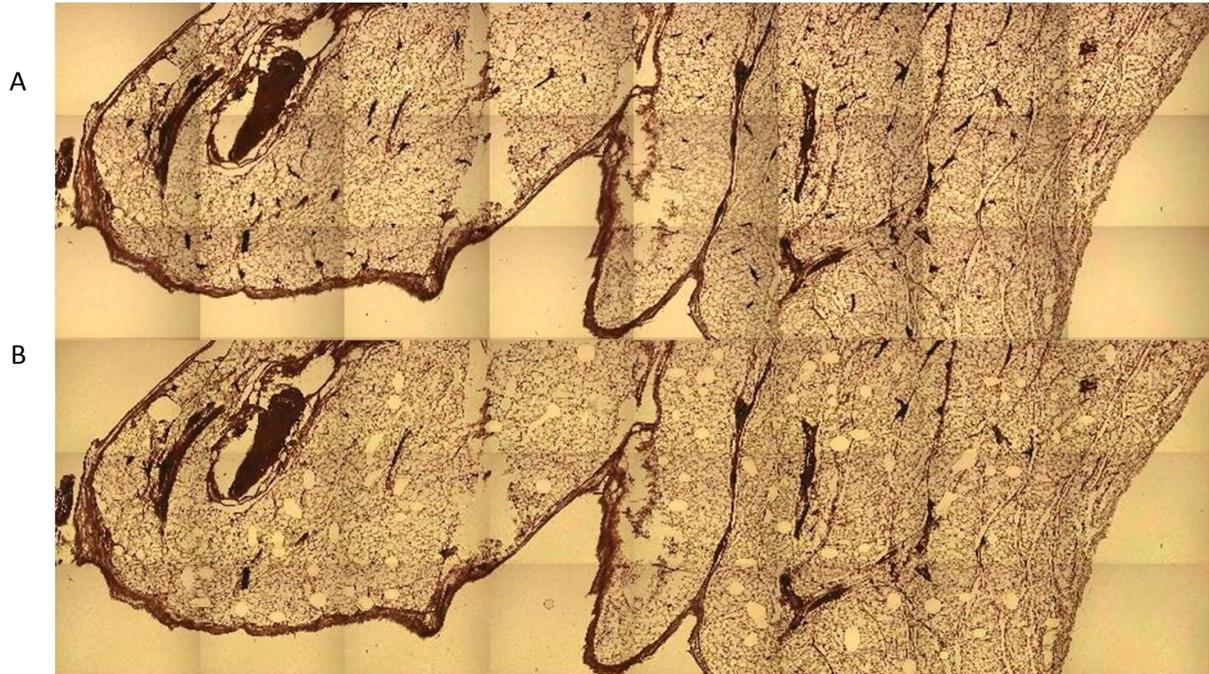


Figure 66: Non-contact laser capture microdissection to enrich for MG epithelial cells.

The overview consist of 24 individual snapshots under 5x magnification. A: MGs were dissected from ACI rats and cryosections were collected on PEN-membrane covered slides, stained with hematoxylin and eosin and viewed to locate the MG epithelium. B: View of the specimen after cutting and catapulting of the target regions.

Methylation data for MG samples containing DCIS was obtained for Msx1, Gbp2, and Npas1. Msx1 and Gbp2 methylation showed less interindividual variation than for Npas1 and resembled methylation of the corresponding bulk MG samples (Figure 68).

In total, with the exception of Thy1 and Zfp507, all epithelial cell samples obtained from LCM mostly confirmed median methylation obtained from the mix of all cell types within the MG samples, although the high variation often complicated concrete comparisons. Thy and Zfp507 clearly exemplified that analysis of a mixture of various cell types, which the MG is comprised of, might affect median methylation levels of a given specimen.

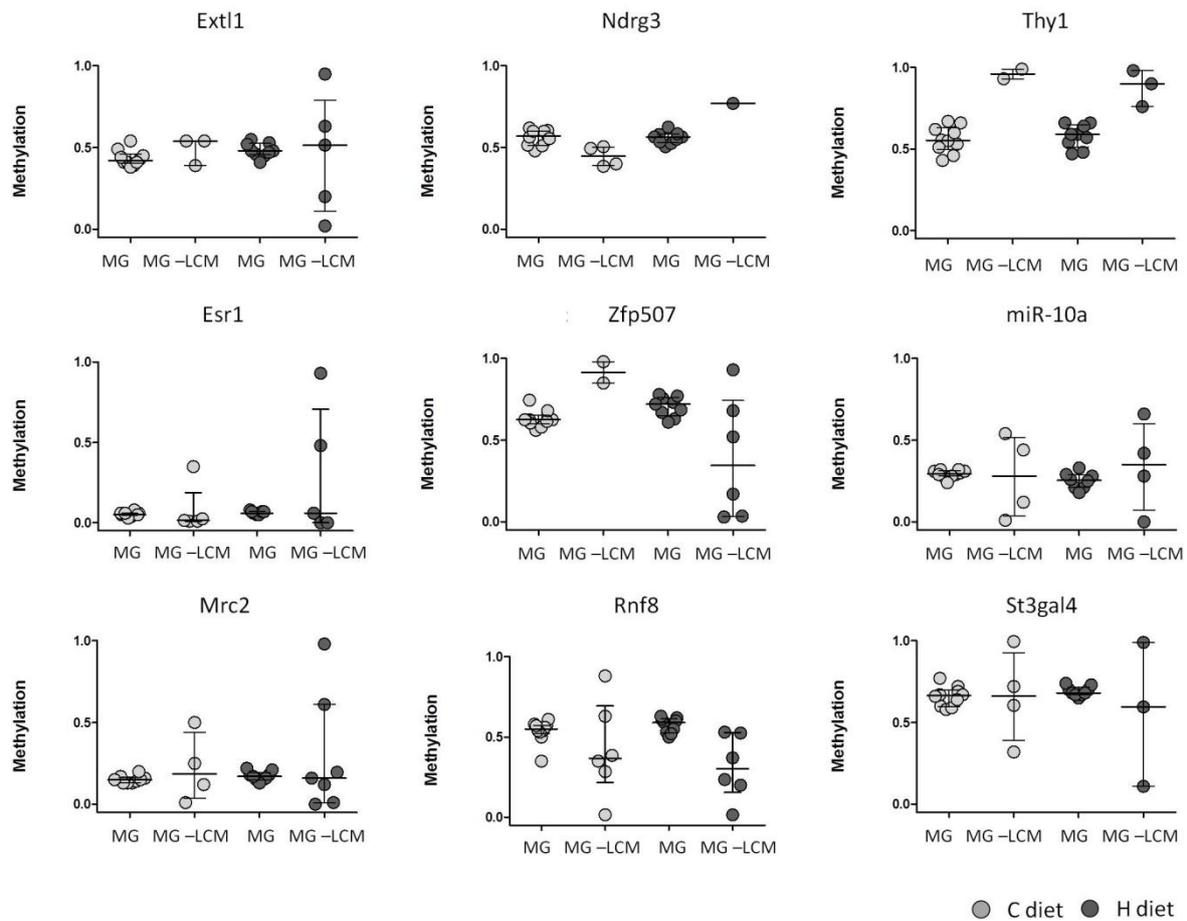


Figure 67: Quantitative analysis of candidate regions of enriched MG epithelial cells.

Median methylation in healthy mammary glands (MGs), enriched mammary gland epithelial cells obtained by LCM (MG-LCM) from PND 240 analyzed by MassARRAY. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, IF diet).

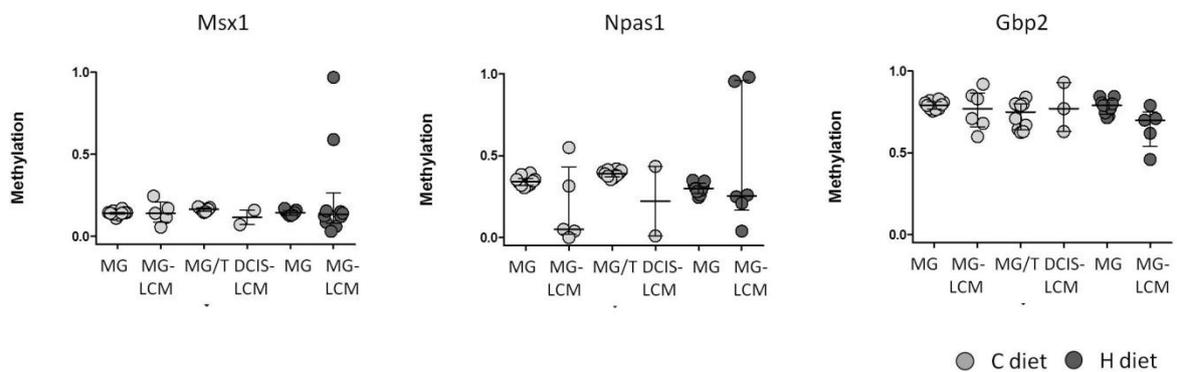


Figure 68: Quantitative analysis of candidate regions of enriched MG epithelial cells with DCIS.

Median methylation in healthy mammary glands (MGs), enriched mammary gland epithelial cells obtained by LCM (MG-LCM), mammary glands with no palpable tumors (MG/T) and ductal carcinoma *in situ* (DCIS) enriched by LCM from PND 240 analyzed by MassARRAY. Each dot in the scatter plots represents one animal (light grey, control diet and dark grey, IF diet).

5. Discussion

Numerous studies indicate that lifestyle and environmental factors, such as dietary habits, essentially affect international incidences of certain cancer types [124, 125]. Notably, hormone-dependent cancer rates are particularly low in populations following a traditional Asian diet rich in soy and soy products. IF have attracted much attention regarding their chemopreventive potential with several investigations providing evidence of IF targeting the epigenome.

In the present thesis, two sequencing-based methodologies for methylation profiling, *i.e.*, MClp-Seq and RRBS were established and implemented to analyze the impact of IF on epigenetic mechanisms in breast cancer prevention. Genome-wide DNA methylation profiles of MGs of healthy rats and during E2-induced mammary carcinogenesis were analyzed by means of bioinformatics evaluation and quantitative confirmation by MassARRAY technology. Additionally, IF-induced epigenetic effects culminated in gene expression changes of candidate genes involved in E2-responsive signaling and epigenetic mechanisms.

Dose-response aspects, critical time windows and the influence on carcinogenesis will be discussed.

5.1 Method establishment

5.1.1 MClp-Seq in Wistar rats

The impact of IF on DNA methylation on a genome-wide scale was investigated using Methyl-CpG Immunoprecipitation (MClp) followed by next generation sequencing (NGS, Seq). Since animals of the MC and MH groups were the first samples available, we selected those for genome-wide DNA methylation profiling.

Affinity enrichment-based methods such as MClp have been used in combination with promoter, CGI, or tiling arrays as well as with NGS. Array-based data analysis was routinely used in the division and provided straightforward bioinformatics analysis, although biased for preselected genomic regions covered by the array. However, technical support was discontinued for the rat genome at the DKFZ, therefore NGS-based detection was applied. We used MClp enrichment for medium and highly methylated DNA fragments by adjusting the amount of salt in the washing buffer, we were able to enrich DNA fragments according to their degree of methylation.

A stepwise elution of fragments from the MBD protein with increasing amounts of salt results in enrichment of fractions with increasing DNA methylation levels and density. By using high salt (HS)

eluted fractions or by applying a single fraction (SF) elution protocol to elute all methylated fractions, high or moderately-to-high methylated DNA fragments, respectively, can be enriched.

5.1.1.1 HS elution: enriching for highly methylated DNA

Firstly, we focused our analysis on the highly methylated regions obtained by the HS protocol. Sequencing the high CpG density regions, we obtained roughly 20 M reads per group and covered 24-28 % of all available CpGs in the rat genome at least once with a good overall genomic coverage (saturation correlation 0.98). Enrichment profiles generated by MChIP-Seq were used to identify DMRs between animals on standard IF-containing diet which were swapped to control diet (MC) or high IF diet (MH) during the hormonal decline period after ovariectomy. MChIP-Seq does not provide precise quantification of methylation levels. Rather, differences between samples are measured as enrichment scores of DNA fragments. The Homer suite of tools was used for analysis of MChIP-Seq data. After bioinformatics comparison, we identified relatively few methylation changes between MC and MH group throughout the genome. With 4-fold difference in read counts between both diet groups, and at least 10 reads per region in one group as selection criteria, we observed an increase of hypo- over hypermethylated regions (3,014 vs. 2,077 DMRs) due to the post-OVX IF intervention.

Annotation of the DMRs according to genomic regions showed that the majority of changes in DNA methylation occurred outside of regulatory regions (promoters, CpG islands or 5' UTRs). In fact, more than 90 % mapped to intronic and intergenic regions, as well as to repetitive sequences (LINE, SINE, LTRs, simple repeats and satellites) introducing changes of DNA methylation within these regions.

The precise function of DNA methylation changes in introns, exons or intergenic regions has not been clarified in detail, but recent investigations suggest that they play a role in transcriptional regulation, alternative splicing and promoter usage as well as in expression of coding and non-coding transcripts [195]. Further, it has been reported that DNA methylation outside of promoter regions might play a role in controlling the distribution of repressive chromatin marks such as H3K27me3 around the border of active core promoters and therefore counteracting their propagation [196]. However, some of the identified DMRs presumably overlap with enhancers and other distal regulatory regions which regulate gene expression in a cell type-specific manner [197]. Different from the human and mouse genome, the exact mapping of enhancers is, however, not well known for the rat genome and will be a focus of future investigations.

5.1.1.2 SF elution: enriching for moderate and highly methylated DNA

Since it has been suggested that highly methylated CpG-dense regions might be resistant to demethylation [181], we were interested in the IF-induced effects on moderately methylated and

probably less CpG-dense regions. Therefore, we performed a second MClp-Seq approach applying the SF elution protocol and sequenced all methylated fractions, including fragments with moderate-to-high methylated DNA and intermediate-to-high CpG density.

Sequencing of medium-to-high methylated DNA fragments after applying SF revealed a number of novel insights about IF-responsive regulatory modules. Comparing HS and SF enrichment techniques both revealed a different representation and coverage of the genome. HS favors high CpG density regions and covered 24-28 % of all available CpGs in the rat genome at least once. 8-9 % of all CpGs showed more than 10x coverage. In contrast, the SF protocol covered 76-85 % of all available CpGs with 23-32 % of all CpGs showed more than 10x coverage.

In total, SF obtained 45-71 M reads per group, which accounts for 2.5 to 4-fold more reads than the HS elution. This difference might be explained by the slightly higher number of animals sequenced. Since these reads were distributed over more CpGs, coverage depth per single CpG was lower than in the HS (saturation correlation 0.92 vs. 0.98). We observed that only few CpGs showed a high read depth with more than 100 reads per site, whereas a high percentage of the CpGs in the SF elution experiment was sequenced to a lower depth which subsequently hampered the identification of significant DMRs. In contrast, the number of highly covered CpGs was higher with the HS elution protocol. These sites are located in CpG-dense regions which were highly enriched and subsequently sequenced to a greater depth, resulting in higher power to detect significant changes.

Bioinformatics comparison of the MC and MH group identified less methylation changes in the SF study than in the HS protocol (1,062 hyper- and 296 hypomethylated DMRs), with a similar relative genomic distribution. However, DMRs showed relatively little overlap between groups, reflecting that both elution protocols cover specific sets of CpG sites due to the differences in genomic coverage.

We required a 4-fold difference in enrichment/reads between two groups to call a region “differentially methylated”. To avoid detection of DMRs with low coverage, we set stringent cutoff criteria of at least 10x coverage per region in one group. Therefore, more than half of the newly covered CpGs using SF elution had to be discarded due to insufficient coverage but with the advantage to detect biological meaningful significant changes.

In future experiments, it will be necessary to increase the number of reads per sample by additional sequencing or by decreasing the number of samples per lane, to successfully cover this broader range of CpG-densities. It should then be considered that the required sequencing effort would increase overall experimental cost. Also, it has to be taken into consideration that enrichment-based methods only provide indirect DNA methylation information which has to be validated by

quantitative methods, *e.g.*, clonal bisulfite sequencing, pyrosequencing or by MassARRAY technology [94, 198].

5.1.1.3 Confirmation of MClp-Seq data

For validation of differential methylated regions observed in the genome-wide methylation analysis by MClp-Seq, quantitative analysis of DNA methylation was performed using the mass spectrometry-based EpiTYPER MassARRAY technology. This technique represents a candidate region approach and is able to provide accurate quantitative methylation data with a technical detection limit of 5 % methylation difference, by quantifying median methylation of CpG units of candidate regions.

We selected 14 IF-induced DMRs of both MClp experiments mainly located in gene promoter regions or close to the TSS to quantify methylation levels in the experimental groups. For most of the candidate regions, we were not able to confirm IF-induced methylation changes observed by MClp-Seq. Median methylation differences of less than 10 % between the MC and MH groups were observed with high interindividual variation within the dietary groups. The high interindividual variation might be attributed to the fact that the Wistar rat strain is an outbred strain, and therefore individual response to ovariectomy and IF intervention might be variable.

Overall, we were able to identify a statistically significant effect of the post OVX high IF diet on DNA methylation in two regions (Gadd45b and Cyp2C6v1).

A possible explanation for the discordance between the results obtained by MClp enrichment and EpiTYPER MassARRAY might be the composition of the MG samples we used. Histologically, lobules and ducts are organized in single layers of luminal epithelial and contractile myoepithelial cells which are surrounded by the ECM, fibroblasts, immune cells, and adipocytes, as well as blood vessels, and others. MClp is an enrichment-based methylome profiling technique. Consequently, differences in read counts between groups might reflect an enrichment of highly methylated DNA derived from a subpopulation of cells. Confirmation of differences by MassARRAY relies on the quantitative analysis of methylation derived from all cells in one sample. Therefore, the methylation differences of a particular subpopulation will be “diluted” by the average methylation levels of the bulk of contributing cell types. Differences in read counts in MClp-Seq experiments might also reflect differences in the glandular structure or the cellular composition of the MG between samples, rather than alterations in methylation levels [75].

Furthermore, it is not known whether affinity-based methods distinguish 5mC and 5hmC which might lead to validation problems, since quantification by MassARRAY relies on bisulfite conversion which does not discriminate these chemical modifications [53].

Additionally, dietary and environmental factors have been shown to induce rather small changes in DNA methylation in non-neoplastic specimen which is in sharp contrast to the massive methylation changes in cancerous samples compared to normal tissue [199], reviewed in [200, 201]. Here the pathological process entirely disrupts the physiology of the cells, and, therefore, it is not unusual to detect large differences in DNA methylation [83].

In our project, the study with the MC and MH groups provided the first DNA samples available for genome-wide profiling. In fact, if soy IF lead to developmental reprogramming during early life by engagement of the epigenomic machinery, reflected in modified methylation profiles, it is reasonable that we did not detected major methylation differences between these two groups. The MCIp technology has been successfully used previously to detect methylation changes in cancer studies [83, 169, 202]. Since it does not provide quantitative data it might be not optimally suited to identify subtle diet-induced methylation changes. Nevertheless, we found interesting regions susceptible to IF-induced alterations in methylation.

In order to allow a more accurate assessment of the potential of IF on DNA methylation in further studies, we decided to establish RRBS for genome-wide methylation profiling, as it avoids the drawbacks of the enrichment-based MCIp-Seq method and enables a quantitative assessment of small-scale methylation differences of a large fraction of the genome at reasonable costs.

5.1.2 RRBS in ACI rats

Similar to the investigations in Wistar rats, we intended to perform genome-wide methylation analyses to investigate the impact of a lifelong dietary IF exposure on DNA methylation changes in MGs of ACI rats. In the kinetics arm of the project, we planned to focus on developmentally critical time windows, whereas in the carcinogenesis arm, we were interested in cancer preventive efficacy during E2-driven mammary carcinogenesis.

In order to overcome the limitations of the enrichment-based MCIp-Seq technology that we had used in the Wistar study, we decided to implement RRBS in the division for our methylome screen in the ACI rat experiments. The advantage of RRBS lies in the fact that this method provides single nucleotide methylation information for about 10 % of all CpGs in a genome, with enhanced coverage of regions, with moderate to high CpG density, including CGIs, promoters and enhancers [88, 203]. Methods other than RRBS that assess DNA methylation on a single nucleotide basis and a genome-wide scale either rely on high bioinformatics and financial effort (*e.g.*, WGBS) or on a reference methylome available on BeadChips arrays (*e.g.*, Illumina 450k).

In cooperation with Christoph Bock (CeMM, Vienna), we have established a RRBS library preparation pipeline. We started to analyze DNA from pre- and postpubertal ACI rats, since the main changes of MG development and differentiation are initiated at puberty [108]. In the second part of the study, we continued with adult MG and tumor samples.

5.1.2.1 Genome-wide methylation screen during puberty

In order to understand the potential of IF on DNA methylation during the early epigenetic programming of the MG, we subjected 24 MG from pre- and postpubertal ACI rats to RRBS library preparation. Sequencing was done on 6 multiplexed libraries on one lane. We roughly covered 2.18 M out of the 24.7 M CpGs in the rat genome, with roughly 1 M CpGs being uniquely aligned per sample. In total, 14-35 M high quality reads were obtained per sample with an alignment rate of 73-76% and a mean bisulfite conversion rate of 98.5 %. This is in line with results that have been published by others [204-207]. To our knowledge, there is no reference rat methylome of the MG available. Several studies investigated DNA methylation in rat on a genome-wide scale but focused on different tissue such as liver, kidney, placenta, or neurological tissues [204, 206, 208]. Methylation patterns have been shown to serve as a specific cellular memory and are therefore highly cell type specific. For example, embryonic stem cells have very low levels of DNA methylation which is increasing by *de novo* methylation determining somatic methylation patterns with up to 80 % global methylation [209]. Therefore, we could not use these available data sets for comparison and to investigate the impact of IF on the MG methylome.

In total, we determined a global DNA methylation of 64 % in healthy pre- and postpubertal MGs of ACI rats. In order to detect CpGs that responded to environmental perturbations, *i.e.*, dietary IF intervention, we first applied filtering criteria in an attempt to increase the reliability of data being analyzed at a given CpG and limit the occurrence of false-positives. We required high sequencing read depth per CpG per sample, filtering out loci with lower than 20x coverage in at least 10 out of the 12 animals per age group, enabling a detection limit of at least 5 % absolute methylation differences. Comparing rats of individual dietary regimes (C vs. H group), we observed modest differences in methylation with less than 10,000 DMS.

In order to gain potential insight into functional and mechanistic characteristics of DMS, we assessed the overlap with various genomic features with respect to 5' UTRs, promoters, and CGIs, as well as introns, exons, intergenic and repeat annotations. In agreement with the results obtained by MClp-Seq in the healthy adolescent Wistar rats, IF-induced changes in DNA methylation predominantly occurred outside of regulatory regions (promoters, CpG islands or 5' UTRs). Similarly, more than 90 % mapped to introns, and intergenic regions, as well as to repetitive sequences (LINE, SINE, LTRs,

simple repeats and satellite regions). However, DMS showed relatively little overlap between groups, indicating a specific set of CpG sites that is significantly affected by IF exposure in each age group.

These results support the observations from the MClp-Seq analysis that IF-induced DNA methylation might regulate gene expression not only by affecting gene promoters but also by modulating distal regulatory regions such as enhancers or non-coding RNAs. This might be achieved by engaging chromatin-modifying enzymes that influence transcriptional regulation and cellular health. Sandovici et al. investigated the Hnf-4 α locus, a TF involved in glucose homeostasis and development of type 2 diabetes and demonstrated the impact of maternal diet on promoter-enhancer regulation during fetal and early postnatal life in pancreatic islets of rats [210]. Low protein intake of the dams resulted in a modest increase in DNA methylation and concomitant changes in histone marks (loss of H3K4me1, gain of H3K9me2). The introduced epigenetic changes mapped outside the core promoter region and decreased the long-range interaction of the distal Hnf-4 α promoter with the enhancer region, leading to profound silencing of the TF and loss of glucose tolerance [210]. Therefore, promoter-enhancer interaction might be a frequent but hitherto underestimated target of nutritional modification such as early dietary IF exposure.

5.1.2.2 Methylome profiling during E2-induced breast carcinogenesis

Additionally, we mined the DNA methylome of ACI rats during E2-driven carcinogenesis of the MG, in order to assess the influence of a lifelong IF exposure on DNA methylation and potential cancer preventive effects.

After we optimized our RRBS library preparation protocol for the adult rat study and sequenced libraries on two lanes (instead of one as in the puberty study), we were able to increase the qualitative as well as quantitative results of the sequencing run. With an increase in the bisulfite conversion rate to 99 % and an alignment rate of 73-76 %, we covered around 2.27 M of the 24.7 M CpGs of the rat genome. In total, the number of aligned reads increased by roughly 2-fold to 45-68 M high quality reads per sample. Concomitantly, distribution of the reads throughout the genome enhanced the extent of uniquely aligned reads from 1 M to 1.5 M. Also, read depth was increased from 30x to roughly 70x mean coverage per individual CpG.

Overall, we determined a global methylation of 53 % in MG and tumor DNA of adult rats independent of the state of disease. This represents a loss of global methylation in course of MG development of 11 % when compared to 64 % methylation in MGs of the puberty study. This is in line with Heyn et al. who proposed age-specific demethylation in healthy normal tissues and were able to determine the biological age solely by means of DNA methylation [211].

Similar to the approach in the puberty study, we applied filtering criteria in an attempt to increase the reliability of the data at a given CpG, responsive to diet and E2-treatment. In sharp contrast to array-based analyses with defined sites at preselected genomic positions, CpG sites in RRBS runs that pass the selection criteria are not necessarily identical in all samples of one experiment, depending on the quality of the sequenced library. Consequently, the number of overlapping sites that meet certain quality criteria such as coverage or false discover rate might vary substantially. Since we wanted to compare several intervention groups within the carcinogenesis study, we required high sequencing read depth per CpG in all samples. Therefore, we filtered out loci with lower than 20x coverage in at least 20 out of the 24 sequenced animals leading to 666,745 overlapping CpGs between all groups.

Looking for significant methylation changes of at least 5 %, we observed an increase of IF-induced methylation changes by 2-fold compared to the puberty study. These DMS represent accumulated changes of the MG, which eventually putting a cell at a heightened risk for damage or might reduce susceptibility towards a certain disease. We identified roughly 24,000 diet-induced DMS in adult rats analyzed in normal MGs (C vs. H group) as well as in tumors (C+E2 vs. H+E2 group). By far the greatest impact on the methylome was seen during tumor formation with roughly 43,000 to 49,000 E2-induced DMS (H vs. H+E2 and C vs. C+E2 groups, respectively).

During tumor development (C vs. C+E2 and H vs. H+E2), we observed a genome-wide loss in methylation with two-thirds of the identified DMS accounting for hypomethylated sites. This finding is supported by previous studies on E2-induced MG tumors in ACI rats that measure global DNA methylation by the use of a HpaII-based radio-labeled [³H]-dCTP extension assay. HpaII cleaves unmethylated CCGG sites and leaves an overhang that is subsequently radio-labeled with [³H]-dCTP. E2-induced a global loss of DNA methylation which was detectable starting from mild to moderate and severe hyperplasia at PND 75, 100, 140 and 180. Due to the methodology used, no information on candidate regions was available [168, 212].

Annotation of the DMS to genomic features with respect to 5' UTRs, promoters, and CGIs, as well as exon/intron, intergenic and repeat annotations revealed that DMS identified during E2-induced tumorigenesis showed a significantly loss of methylation in repetitive elements. Methylation in these regions contributes to limiting DNA accessibility, and transcriptional activation of repetitive elements has been associated with chromosomal instability [213]. Thus, increasing hypomethylation weakens the defense mechanisms against transposition of mobile elements. A recent study has shown that one endogenous retrovirus (ERV) class, *i.e.*, ERV-K-type family contributes to genome variability in inbred rat strains. Therefore, such a defense mechanism is important for rodents who harbor active ERV within their genomes [214].

Although, RRBS tends to cover regions with high CpG density such as promoters and CGIs, less than 15 % of the identified DMS in our data mapped to regulatory regions (5' UTRs, promoter, CGIs). This is comparable to recently published RRBS data sets with less than 20 % of the DMS cover promoter regions [206]. Since we used the Homer suite of tools for annotating the DMS, identified promoter regions are defined as 1 kb around the TSS by default. Standard promoter definition consider 2 kb upstream and 0.5-2 kb downstream of a TSS; therefore, many putative promoter CpGs might have been considered to map to intergenic regions or to exons and introns, explaining the lower number of mappable CpGs at promoter regions.

Notably, we observed a significant enrichment of hypermethylated DMRs in introns, and exons. as well as in regions close to TSS which, in addition to potential roles for general transcriptional regulation, could indicate effects on gene splicing and isoform-specific regulation. Sati et al. observed that specifically intron-exon-intron junction are marked by DNA methylation in rat liver, emphasizing the role of DNA methylation in distinguishing exons from introns [215]. There is also evidence that gene body methylation, particularly in CGIs, could highlight non-canonical TSS and tissue-specific gene regulation as well as dynamically target alternative splicing of cassette exons (skipping or inclusion) and intron retention in the final transcript [216, 217].

DNA methylation and consequently transcriptional regulation of imprinted genes is often affected during cancer development, since these regions require tight control to maintain paternal- or maternal-specific allelic-transcription [218]. So far, we have not investigated these regions in greater detail. Reports by others revealed an impact of endocrine disrupting agents, *e.g.*, TCDD, on DNA methylation status of imprinted genes of embryos at the preimplantation period, influencing the expression of the imprinted genes and fetal development [219].

5.1.2.3 Stringent cutoff criteria for RRBS data secure reliable quantitative methylation data

RRBS provided us a hands-on tool to explore the epigenetic effects of IF on MG development and the risk of MG tumor formation at single CpG site resolution. RRBS enables a direct quantitative readout of methylation levels between samples, theoretically, does not require independent validation.

According to our knowledge, no study so far attempted a quantitative confirmation of RRBS-derived methylation data for validation of identified DMS using MassARRAY technology as second quantitative method.

Results obtained in the puberty study

In a first technical validation, ten candidate regions from the RRBS data of pre- and postpubertal ACI rats were selected for MassARRAY validation, using the same DNA samples that were used for the

RRBS run, to demonstrate the accuracy and reliability of the sequencing results. Each differentially methylated region (DMR) was represented by at least 2 single CpGs or a CpG unit consisting of 2 CpGs. Six regions showed high absolute methylation >50 % and IF induced hypomethylation, whereas four regions with methylation levels below 50 % gained methylation in the animals receiving IF diet. Differential methylation detected by RRBS between C and H diet groups could not be confirmed by MassARRAY. Group wise methylation levels were much more homogeneous when determined by MassARRAY indicating that the methylation differences detected by RRBS were at least partly due to technical artifacts.

Bisulfite treatment of the DNA might introduce a bias towards increased methylation levels since unmethylated non converted Cs will be considered as methylated Cs. PCR amplification errors have also been observed for bisulfite converted DNA (unpublished results Gerhäuser lab). Since RRBS libraries are prepared from low input material (<100 ng) and bisulfite conversion commonly destroys >99 % of intact DNA fragments due to the denaturing environment [220], sequence specific amplifications of preferentially un- or fully methylated DNA fragments have been reported (oral communication PD Dr. Dieter Weichenhan).

However, although diet-induced methylation differences could not be confirmed by MassARRAY technology, in general, the methylation levels detected by both technologies were comparable and showed highly significant correlation (Spearman correlation coefficient ρ : 0.7038).

Results obtained in the carcinogenesis study

For the second technical validation, eleven candidate regions from the RRBS data of the carcinogenesis experiment were selected for MassARRAY validation, using the same DNA samples that were used for the RRBS run. In order to reduce the number of low-confidence CpGs, we integrated additional filters to our selection pipeline to remove sites with high within-group variance. This was intended to limit the inclusion of sites with potential false (low or high) methylation which could contribute to inflated differences between groups in our data set. In a previous study, Watson et al. calculated outliers as maximal observed CpG value minus minimal observed CpG value. If the difference was >30 % CpGs were discarded. In their RRBS data set, bad quality CpGs represented one quarter of all high quality reads with more than 30x coverage per CpG (roughly, 208,000 of 776,000 CpGs) [206]. Therefore, outlier removal with standard deviations exceeding >25 % absolute methylation is highly recommend. Taking this into consideration, we combined settings which best identify CpG sites that could be confirmed by alternative methods. We narrowed down the DMS list by increasing the methylation difference to >15 % and by stringently filtering for standard deviation <15 % absolute methylation to remove sites that were identified as DMS as a result of individual outlier values. Further, we focused on CpG sites with mean methylation levels between 5-95 % to

exclude sites which were either unmethylated or fully methylated, a signature which might be consistent with underlying genetic polymorphism. We were not able to confirm these large interindividual variations by MassARRAY analysis in the previous validation experiment as described above. Therefore, sequencing or bioinformatics artifacts, indeed, most likely account for these observations and it was reasonable to exclude these sites.

Overall, by applying these optimized settings, we were able to confirm tumor-induced methylation changes as well as preventive effects by IF intervention. Due to the low number of identified DMS, we did not reduce the interindividual variation within one group by setting cutoff limits for standard deviation or overall mean methylation for the comparison of C vs. H groups. Consequently, we were not able to verify the IF effects in healthy MGs. However, overall, we demonstrated the consistency of the RRBS and MassARRAY data and calculated significant correlation (Spearman correlation coefficient ρ : 0.7868).

Although methylation differences at single CpGs have been demonstrated to influence gene expression patterns [221], it can be difficult to assess the significance and reliability of variation observed on a single CpG basis, particularly from a statistical perspective, when implementing multiple testing. Since we and others have shown that methylation levels are strongly correlated at CpGs that are in close proximity, methods that assess concordant changes in DNA methylation at neighboring CpGs would circumvent this issue. Thus, defining and analyzing DMRs with 2-3 significantly concordant DMS (either hyper- or hypomethylated) would increase the reliability of the data analyzed and would further enhance the statistical power [222].

Additionally, RnBeads analyses have been based on differences in the group methylation mean values which is highly susceptible to artifacts due to outlier values. Calculations on group wise median methylation values might provide a more reliable tool to call DMS between groups. This option is now implemented in RnBeads.

In order to improve the technical accuracy of RRBS and to identify reliable DNA methylation patterns, unique molecular identifiers (UMIs) can be used to control for amplification biases [223, 224]. These short nucleotide sequences can be tagged to the MspI digested fragments before bisulfite conversion. After PCR amplification, UMI-labeled DNA fragments are digested by enzymatic restriction. Libraries are prepared according to standard RRBS protocols, sequenced, and mapped to the genome. Since the same UMI is attached to different patterns and is therefore randomly distributed throughout the MspI fragments, non-uniquely labeled DNA fragments covering a CpG are considered as PCR-introduced duplicates. The numbers of each fragment can increase tremendously, when using low input material and high cycle numbers during amplification, yielding to a total

duplication rate up to 50 % [224] which can further increase when FFPE material or single cells are used (oral communication David Brocks).

5.2 Biological effects

5.2.1 Investigations on IF dose and exposure time in Wistar rats

As discussed in section 5.1.1 we performed a genome-wide methylation analysis using MChp-Seq to identify genomic regions differentially methylated between the MC and MH diet groups. These were exposed to intermediate levels of IF until OVX and differed only in IF exposure during the hormonal decline after OVX. Overall, we detected only few DMRs between these two groups, which might not be surprising, as the rats were similarly exposed to IF during important developmental periods (*in utero* until adulthood). In comparison to this exposure period, the post OVX period, when IF exposure of the rats was either stopped or even increased, was relatively short. Migration studies have demonstrated that environmental and life-style factors (*e.g.*, diet) are able to affect susceptibility to specific types of cancers but might take several generations to occur [225].

We selected genomic regions for validation of the DMRs by MassARRAY. Although methylation differences between the MC and MH group could not be confirmed, quantitative methylation analyses of additional dietary groups revealed interesting time- and dose-dependent influences of IF exposure on DNA methylation and gene expression, as discussed in detail below.

5.2.1.1 Specific TF are enriched in DMRs identified by genome-wide analyses

In order to investigate whether IF-induced DMRs share common genomic features, a *de novo* transcription factor (TF) binding motif discovery as well as an enrichment analysis of known TF binding motifs was performed using the Homer suite of tools [172]. These motif analyses enabled a deeper insight into the molecular mechanisms affected by the IF-induced DMRs identified between MC and MH by MChp-Seq.

Distinct results were obtained when analyzing the datasets of the high salt (HS) vs. single fraction (SF) elution for motif enrichment. Only little overlap between enriched motifs was detectable between HS and SF protocols. Enrichment was predominantly found for basic leucine zipper (bZIP)-domain TFs (*i.e.*, Atf3, Fosl2 and MafK) in hypomethylated DMRs. This family of TF plays important roles in cellular processes including cell proliferation, differentiation, and apoptosis and has been implicated in malignant transformation. The AP1 FOS/JUN heterodimer is a well-known member of this family which is activated by growth factors such as EGF, TGF α , IGF and E2. ERs also mediate gene transcription *via* AP1 binding sites particularly when transactivation properties are modified by the

ligand bound, *e.g.*, tamoxifen and raloxifen with ER β being 10-fold more efficient in activating AP1 than ER α [226]. However, ER-stimulated transcription of AP1 targets is associated with an induction of cell proliferation and is regulated by cyclin D1 and E2F [227]. Thus, hypomethylation of regions with binding sites for bZIP-domain containing TF which potentially dimerize with JUN to form an active AP1 might reflect a mechanism by which the additional high IF exposure after OVX provokes the increased proliferative and estrogenic responses indicated by increased PR and PCNA IHC staining in MH group animals vs. the MC group.

Motifs of E26 transformation-specific (ETS)-domain TF family members (*i.e.*, ERG, SpiB and Ets1) were enriched in hypermethylated DMRs. ETS-domain proteins have been implicated to play important roles in promotion of differentiation, inhibition of apoptosis and in regulation of cellular senescence. Deregulation by inappropriate expression or expression as fusion proteins has been associated with the development of certain types of cancers, *e.g.*, ovarian or prostate cancer [228]. Growth factor dependent phosphorylation *via* the proto-oncoprotein Ras and subsequent activation of the MAPK (mitogen-activated protein kinase) pathways has been tightly connected to ETS-domain containing TFs [229]. There is evidence that the transcriptional-regulation process also involves epigenetic modification of the targeted promoters, *e.g.*, Ets-1 is able to bind to CBP/p300 and enhances transcriptional activation of downstream genes [229]. Therefore, hypermethylation of ETS-domain TF binding sites might provide a functional mechanism by which IF put their growth limiting and beneficial health effects into action.

The finding that hypo- and hypermethylated regions are associated with functionally distinct TF motifs suggests that hypo- and hypermethylation may occur in distinct cell types. It is intriguing to speculate that there might be a common mechanisms targeting methylation changes to specific sites and alterations in distinct cell types might converge on pathways, regulating the ability of the MG to respond to IF exposure.

Overall, these data provide evidence that IF affect gene categories by modulating DNA methylation and co-operative binding of transcription regulatory factors which consequently might affect the susceptibility of the MG to pathological processes.

5.2.1.2 IF dose aspects

One aim of the first project analyzing MGs of Wistar rats was to investigate the impact of two different doses of lifelong IF exposure on DNA methylation and estrogen responsiveness. A three-day uterotrophic assay was conducted to characterize the general physiological response and E2-modulating ability of the different IF exposure scenarios.

IF doses were chosen based on reported dietary IF intake of Western and Asian populations [230]. The control diet contained very low IF amounts (3 mg/kg diet) yielding to a total intake of the animals of 0.3 mg/kg bw/d that resembles the dietary habits of the Western population with a daily intake of 1-2 mg IF aglycone equivalents.

Taking into account the IF exposure in Asians and the average daily food uptake of rats, we performed an allometric conversion to apply diets with different IF concentrations [231, 232]. Two experimental animal chows were enriched with the NovaSoy650® extract to a final concentration of 68 mg/kg diet in low (L) and 503 mg/kg diet aglycone equivalents in high (H) IF-enriched diets. The medium IF concentration diet (M) was a standard soy containing diet provided by Ssniff GmbH which contains soybean meal and full fat soybeans as protein and fat source, resulting in a final concentration of 232 mg/kg IF aglycone equivalents.

Based on the allometric conversion (considering women with body weights of 60 kg), the L diet mimics an IF intake of 0.89 mg/kg bw and would imply a daily IF uptake of 53 mg. Regarding the average daily dietary IF intake of Japanese adults of 25-50 mg aglycone equivalents the L diet resembles the average Asian dietary behavior. Similarly, the M diet represented an IF intake of 2.96 mg/kg bw, which can still be achieved by dietary soy consumption. A daily IF intake of 177 mg would represent approximately 10 % of the Asian population which ingest more than 100 mg/day [230]. In contrast, the H diet leads to a higher IF dose of about 5 mg/kg bw, which would imply a daily IF uptake of 390 mg. This represents a very high IF intake which might not be relevant as a dietary exposure scenario. However, women undergoing self-medication using IF-containing capsules as HRT can achieve IF intakes greater than 200 mg/d. Products on the market contain between 40-150 mg IF/capsule, with a daily recommended dose of 2 capsules. Thus, elderly postmenopausal women might increase their IF intake far beyond what is observed for Asian women [7, 233]. Additionally, soy protein-based infant formulas represent a significant source of IF. Dinsdale et al. suggest that approximately 37-43 % of North American infants are formula fed from three to six months postpartum, thereby achieving IF doses of 5.7-11.9 mg/kg bw during the first months of life [234].

Onset of puberty is accelerated by lifelong exposure to low and high IF doses

In order to assess the estrogenic properties of a low (L) and high (H) dose lifelong IF intervention we investigated alterations of puberty onset and uterus wet weight.

Interestingly, time to onset of puberty was reduced by both dietary regimes when compared to the control group. This is in line with previously published results, where administration of dietary IF in a multigenerational setting with different doses accelerated vaginal opening [235-237]. Also

investigations in humans that have examined markers of sexual development demonstrated significant effects of early IF exposure. Adgent et al. investigated 2920 white US American girls and identified a 4 month earlier onset of menarche among soy-based fed girls when compared to girls fed with non-soy-based infant formula [238]. Additionally, Kim et al. reviewed a number of studies conducted in US, UK, German and Korean populations and reported a positive association of premature thelarche (MG development) and/or menarche after early exposure to IF [239]. Since early onset of menarche is considered as a risk factor for developing MG tumors in humans these observations have to be interpreted as a potential adverse effect and hold a substantial public health relevance [240]. However, there was also conflicting data showing a deceleration of pubertal timing in response to dietary IF. Cheng et al. investigated 227 girls from the Dortmund Nutritional and Anthropometric Longitudinally designed study and reported a delayed Tanner stage 2 for breast development in girls of highest compared to lowest dietary IF tertile (1199 vs. 13 $\mu\text{g}/\text{d}$) [241]. Similarly, D'Aloisio et al. investigated 50,884 US and Puerto Rican women and identified associations for both, very early (<10 years) and late (>15 years) onset of menarche with postnatal soy-based formula nutrition [242]. Menarche is primarily controlled by the hypothalamic-pituitary-ovarian axis while thelarche and pubic hair development is modulated by hypothalamic-pituitary-adrenal axis. Therefore, IF exposure as well as several other factors, *e.g.*, body mass, and body composition, as well as protein intake, and protein source might influence distinct pubertal events in a different manner [243].

Treatment with E2 for three days resulted in an increase of the UWW, whereas exposure to IF by the different diets exerted no effects on UWW. This is in line with a previous report from Molzberger et al. using an IF enriched diet containing 300 ppm aglycone equivalents [235] and argues against a direct estrogenic effect of IF exposure.

Alternatively, impaired estrous cyclicity and ovarian function have been reported as a function of a complex influence of IF on the previously mentioned hypothalamic-pituitary-ovarian axis [239]. In the peripubertal period when endogenous E2 levels are low, GEN exposure induced increased sensitivity of the pituitary gland to gonadotropin-releasing hormone (GnRH), altering GnRH-induced luteinizing hormone (LH) release. This response seems to be dose-dependent. In CD rats, subcutaneous injections between PND 15-25 \pm 2 with low dose (10 $\mu\text{g}/\text{d}$) GEN increased and high dose (\geq 100 $\mu\text{g}/\text{d}$) GEN decreased LH release, resulting in altered time of ovulation [244].

These data indicate that indirect regulatory activities of IF also have to be taken into consideration when interpreting IF-mediated physiological effects.

Lifelong high but not low IF doses reduce proliferative and estrogenic response

We investigated the impact of IF on expression of PR as the prototypical marker for E2 responsiveness in the MGs by IHC. E2 exposure stimulated PR protein expression, whereas this estrogenic response was clearly reduced by lifelong exposure to the H diet. In contrast, the IF dose in the L-diet group was not sufficient to reduce estrogen responsiveness. Interestingly, the intermediate IF level of the MC diet (but not of the MH diet) lowered the sensitivity of MGs towards E2 treatment.

We further investigated the proliferation marker PCNA in the MGs by IHC. Similar to PR protein expression, E2 exposure stimulated PCNA protein expression. E2 response was clearly reduced by exposure to high as well as medium (MC, but not of MH diet) IF-containing diets, while the low lifelong IF dose was not sufficient to lower sensitivity of MGs towards E2 treatment. This is in line with earlier observations, where effects on differentiation of TEBs, highly fragile structures in the developing MG, were only achieved using higher doses of dietary GEN [245]. In contrast, in the absence of additional E2 stimulation a low dose of IF (L as well as MH diet) weakly induced the percentage of PCNA-labeled nuclei and increased mRNA transcript level of Ki67 (but not PCNA).

Protein expression determined by IHC of the selected molecular markers (PR, PCNA) did not correlate well with mRNA expression. Pronounced E2-mediated induction of mRNA expression was only detectable for Pgr as well as for Greb1, a second E2-responsive gene investigated. Additionally, prevention of E2 stimulated mRNA induction was rather seen in the low instead of the high exposure group. While protein expression can be evaluated by IHC in the cells of interest, for example, MG epithelial cells, determination of mRNA expression is often performed with bulk tissue where additional cell types, *e.g.*, adipocytes, fibroblasts, and lymphocytes contribute to mRNA levels. Additionally, these discrepancies might be due to differences in the kinetics of induction and the half-life of mRNA and protein. The proliferation markers differ in their mRNA stability with PCNA showing faster mRNA degradation and stronger protein stability [246].

In summary, reduced proliferative response and reduced estrogenic sensitivity could only be observed in the animals exposed to medium and high IF doses. However, pro-estrogenic effects were particularly observed in the low dose group, *i.e.*, induction of proliferative response without E2 challenge as well as accelerated onset of puberty.

Lifelong low IF exposure induces a U-shape like DNA methylation and mRNA expression profile

In order to assess whether different IF doses are able to induce alterations in DNA methylation, we selected 14 DMRs identified between the MC and the MH group from both MCIp experiments for quantitative methylation analyses by MassARRAY technology.

First, we looked at all groups exposed to high IF levels post OVX and compared the trends of the different IF doses administered before OVX (*i.e.*, no, intermediate, or high doses in groups CH, MH and HH). With increasing dietary IF concentration to medium and high levels pre OVX, methylation often increased with a linear tendency, whereas short-term exposure to high IF levels only after OVX (CH) largely lowered methylation levels relative to control.

Next, we were interested whether a lifelong low dose IF treatment (group LL) would be sufficient to induce the effects obtained with medium (M) and high (H) IF diets. We compared the groups LL and HH, two treatment regimens with a constant lifelong IF exposure, to group CC with no IF exposure. Interestingly, lower IF levels exhibited opposite effects when compared to CC and HH groups and introduced pronounced hypomethylation in MGs of rats leading to a U-shape like dose-dependent DNA methylation and mRNA expression profile. In five out of 14 genes investigated different lifelong IF exposure dose-dependently induced alterations in gene expression *i.e.*, Arpp21, PTEN, Cldn4, Esr1 and Extl.

Arpp21 (cAMP-regulated phosphoprotein 21) is located at chromosome 3p where loss of heterozygosity has been reported in 87 % of primary breast cancers. Embedded in the intron of the Arpp21 gene miR-128-2 is frequently downregulated along with Arpp21 in breast cancer, where it is negatively associated with clinico-pathologic characteristics and survival outcome [247]. Induced gene expression due to a high lifelong IF exposure would provide a clear benefit for individuals at high risk for breast cancer. PTEN (phosphatase and tensin homolog) is one of the most frequently mutated TSG. It negatively regulates the PI3K/AKT pathway that promotes cell proliferation, growth and motility and alters the functions of several AKT substrates, *e.g.*, Mdm2, a proto-oncogene (E3 ubiquitin protein ligase) that targets tumor suppressor proteins such as p53 for proteasomal degradation and BAD, a BCL2 antagonist of cell death [248]. Here, particularly the low IF diet increased gene expression representing a clear beneficial effect. Cldn4 (claudin 4) plays a crucial role in the formation and maintenance of tight junctions and is involved in breast cancer progression. It has been shown that Cldn4 overexpression increased cell proliferation and migration in MCF-7 cells transplanted into nude mice and gene expression was regulated by DNA methylation [249]. Since the disruption of tight junctions can lead to the loss of intercellular cohesion, downregulation of Cldn4 by lifelong high IF might contribute to reduced cell invasiveness and increased differentiation of cancer cells [250]. Esr1 (estrogen receptor 1) encodes the ER α which plays a significant role in the development of breast cancer. One-third of all breast cancers lack ER α and rarely respond to hormonal therapies. They show higher proliferation and a worse clinical outcome than ER α -positive breast cancers [251]. Lifelong high IF exposure reduced Esr1 as well as Extl1 expression (exostosin-like glycosyl-transferase 1) which was identified as one of the top E2-inducible genes in rat MGs

[173]. However, its role in human breast carcinogenesis has not been investigated yet. A reduction of *Esr1* mRNA levels in an ER α -positive MG might not be interpreted as a negative effect though since it might solely reflect a tissue-specific autoregulation of the ER α , as previously observed after exogenous E2 treatment [252].

Since we selected particularly IF affected DMRs for quantitative analysis, we did not see pronounced E2-induced DNA methylation or gene expression changes. Solely for *Extl1*, exposure to high IF levels increased the susceptibility to estrogen, but only when administered in a lifelong fashion.

Overall, there seems to be a critical IF concentration leading to distinct biological effects. After exposure to the lower IF concentration, we observed opposing effects to those seen with medium or high IF levels mirroring maximum Asian exposure. These ‘hormetic’ or ‘biphasic’ dose-response models have been perceived as the pharmaco-dynamic nature of a number of substances particularly in tumor drug development [253, 254]. Therefore, the impact and efficacy of IF might be dependent on two situations; either low-dose stimulations lead to adverse effects (*e.g.*, increased E2-sensitivity of the MGs, induced cell proliferation) or define the therapeutic zone and show beneficial effects (*e.g.*, induced expression of TSGs) [255]. However, we could not always observe a clear dose-response pattern.

Interestingly, early studies already determined these biphasic effects of IF. Incubations with GEN in MCF-7 showed that GEN concentrations below the IC₅₀ (10-50 μ M) stimulated tumor cell growth in ER α -positive cells, while a higher GEN concentration lead to an inhibition of genes involved in cell cycle control [256]. Similar observations have been made for tamoxifen which induced prolactin secretion in rats when administered at low concentrations (100 nM) but not at higher levels [256]). Further, Iwasaki et al. investigated 24,226 pre- and postmenopausal women in the Japan Public Health Center-based prospective study and detected a similar biphasic effect regarding IF intake and breast cancer risk. There was an inverse association between plasma GEN levels and the risk of breast cancer since adjusted odds ratio (OR) gradually decreased with increasing GEN plasma levels up to 0.34 for the 4th quartile. Interestingly, OR were reduced to 0.79 in the 4th quartile by increasing DAI levels but were adversely affected by lower DAI levels as in the 2nd (1.50) and 3rd (1.44) quartile (not significant) [257].

5.2.1.3 Critical time windows

The timing of exposure seems to be the most critical factor. There is an ongoing debate whether IF need to be ingested for a lifelong period or if specific critical time windows are sufficient to induce the benefits associated with IF intervention. In order to gain a comprehensive insight into the action

of IF, we covered major developmental windows which have a key impact on MG development and health.

Several earlier investigations have shown that especially a multigenerational and prepubertal exposure to IF modulate MG morphology and provide anti-tumorigenic effects [141, 245]. Several groups of rats in our study were kept on lifelong IF diets containing three different IF doses (L, M and H), which resembles the situation of the Asian population. Here, IF exposure starts already *in utero*, since IF are able to pass to placental barrier and accumulate to quantifiable levels in the amniotic fluid [258]. It is considered that the developmental effects attributable to chronic IF exposure in multi-generational studies result from this fetal exposure, since lactational transfer of GEN is limited as reported for Sprague Dawley rats [259].

We focused on two exposure periods compared to a lifelong exposure, *i.e.*, exposure during puberty only, when the MG is developing, and exposure after OVX, mimicking postmenopausal exposure in humans, for example, for HRT.

Exposure during puberty

Beside pregnancy, the pubertal period is the major period of postnatal MG development and represents a critical time window for epigenetic reprogramming [144]. It is of great clinic-therapeutic interest whether IF exposure during puberty would be sufficient to introduce the beneficial effects observed for a lifelong dietary IF intake. Since daily application of high IF doses during this developmental window of susceptibility would be easier achievable in Western populations than lifelong exposure, this would represent a great achievement for primary chemoprevention approaches of breast cancer.

In order to investigate whether dietary IF exposure only during puberty modulates the MG methylome, a group of rats was assigned to a non IF-containing diet before and after puberty and were exposed to the high level IF diet only during the pubertal period (group CHC).

Interestingly, administration of dietary IF only during the pubertal phase accelerated the time to vaginal opening similar to lifelong IF exposure which is in line with previously published results [235]. In comparison to the control group, pubertal IF exposure prevented the E2-induced gain in UWW after whereas lifelong IF exposure sensitized the uterine tissue to the E2 treatment. It would be interesting to determine expression of ER α and ER β in uterine tissue under these exposure scenarios. Differential ER expression after pubertal vs. lifelong exposure might explain the differential response to E2.

In our study, pubertal H exposure (CHC) prevented E2-induced Ki67 expression and increased E2-sensitivity towards enhanced PCNA transcript levels. This was different from the results of

Molzberger et al. [235]. The physiological relevance of these findings needs to be further investigated. Expression of estrogen responsive genes was not affected by pubertal IF administration, indicating a lack of direct estrogenic effects. On the other hand, Vdr (vitamin D receptor), a negative growth regulator of breast cancer cells and the TF Gata3 were induced in the CHC group, whereas lifelong exposure reduced the expression of both genes. Wietzke et al. reported that IF-induced upregulation of Vdr may sensitize cancer cells to the effects of vitamin D(3) analogs since 80 % of human breast tumors express Vdr at low levels. Therefore, a high IF exposure during puberty may provide a concept to modulate the MG epithelium towards higher cellular vitamin D(3) sensitivity and eventually to reduced cell growth [260]. Gata3 is involved in E2 signaling and an important regulator for luminal differentiation. In a study by Thomsen et al. Gata3 was one of the top inducible genes by prepubertal IF exposure (270 mg Prevastein (46.19 % IF)/kg diet), and lactational E2 treatment. Subsequently, the presence of luminal epithelial cells in the MGs increased which reflects an enhanced glandular differentiation, a phenomenon which can be associated with reduced breast cancer risk [261].

Exposure to the high IF dose exclusively during the pubertal period exhibited only minor effects on DNA methylation levels of the investigated candidate regions and did not seem to be responsible for the effects introduced by a lifelong IF exposure, as we observed opposing effects between the CHC and HH groups. Median DNA methylation of *Extl1* was significantly reduced in the CHC group compared to the control, consistently with or without additional E2 exposure. Hypomethylation of *Cldn4*, which plays a crucial role in the formation and maintenance of tight junctions and is involved in breast cancer progression as well as *Aldh1L1*, a gene involved in one carbon metabolism and relevant for providing SAM for methylation reactions, were detected only in the E2 group (not significant for *Cldn4*). In contrast, we observed reduced methylation levels of *Niacr1* only without additional E2 exposure. The niacin receptor 1 is a G protein-coupled receptor located predominantly on adipocytes that mediate anti-lipolytic effects. It is activated by the endogenous ligands β -hydroxybutyrate and lactate, two metabolites produced during prolonged starvation and intense exercise, respectively. Regulation of expression of this receptor is crucial for whole body energy homeostasis and is highly influenced by high fat diet [262]. However, pubertal IF exposure did not significantly affect the expression of any of these genes analyzed for methylation changes.

Differential effects between short-term pubertal and lifelong IF exposure might reflect direct (anti-)estrogenic properties of IF vs. an indirect influence on gene expression *via* the hypothalamic-pituitary-gonadal axis as described above. In addition, Molzberger et al. demonstrated differential (microbial) metabolite levels between pubertal and lifelong IF exposed animals, with lower levels in

animals that were only exposed during puberty. This might be due to differences in the expression of genes involved in IF metabolism or changes in the gut microflora [235].

Overall, our analyses revealed only few significant changes in gene expression or DNA methylation after pubertal exposure to IF. These results indicate that the pubertal period might not be the most critical time frame for the long term preventive effects of IF.

In addition to the pubertal exposure, several earlier studies have investigated the impact of a pre-pubertal IF exposure. Warri et al. reviewed several reports on various IF exposure scenarios and obtained a conflicting dataset for the pre- and peripubertal periods [245]. Thompson et al. investigated the effects of lactational and postweaning IF exposure in mice in comparison to E2 administered during the lactation period. Interestingly, MG morphogenesis was significantly modified by all treatment regimens towards an increased number of branches of the ductal tree which depicts an early growth stimulatory effect of the prepubertal intervention. Further, the number of highly proliferative TEBs was increased at PND 28 and consequently reduced at PND 42-43 due to the E2 and postweaning IF exposure, reflecting an acceleration of the MG differentiation potentially resulting in a reduced cancer risk [261]. In a study of Lamartiniere et al., prepubertal IF exposure accelerated differentiation of the epithelial cells by transient upregulation of EGF-signaling and protected against chemically induced cancer in rats later in life [263]. Wang et al. identified changes in the proteome of female offspring that was exposed prepubertally to GEN (250 mg GEN/kg diet) *via* the lactating dams. Using two-dimensional gel electrophoresis and mass spectrometry, more than 20 proteins were detected as differentially expressed due to the lactation transmitted GEN compared to unexposed animals, including vascular endothelial growth factor receptor 2 (VEGF-R2) and EGFR [264].

Collectively, these studies indicate that the MG might be more sensitive to IF exposure during the prepubertal phase than during puberty. In an ongoing study on DNA methylation in rat MGs, this aspect is currently taken into consideration.

Exposure post OVX mimicking menopause

IF-containing products are advertised extensively as a natural alternative for HRT, since they potentially reduce menopausal symptoms without adverse effects regarding increased risk for cardiovascular events or breast cancer. Ingestion of the products on the market can yield to high daily IF intakes (40 – 150 mg IF/ capsule). However, high dose IF exposure in postmenopausal women might induce a different response in the MG than lifelong intervention during all susceptible periods and might affect the MG in an adverse manner. This mechanism is not well understood so far.

In order to mimic the human postmenopausal situation, animals received a high IF dose (CH) after ovariectomy during the hormonal decline period. This group was compared to the control group without IF exposure and the group with lifelong high IF levels. These exposure scenarios resemble the Western and Asian population, respectively with a very low (<1 mg) and a rather high (>200 mg) daily premenopausal dietary IF intake. Here the main focus was whether a high dose IF exposure after menopause (ovariectomy) might induce a contraindicative pro-proliferative effect on the MG epithelium.

As in the other parts of the project, we first analyzed the influence of the IF diets on uterine response, alone and in combination with E2 challenge. Interestingly, without E2 challenge the high dose IF exposure after OVX did not exert an estrogenic effect on UWW, whereas the same diet further sensitized the uterus towards E2 treatment. Detection of reference gene expression still indicated a weak estrogenic effect of the post OVX IF diet. mRNA expression of Pgr and C3 was induced with and without additional E2 stimulation. Also, transcript levels of PCNA were increased, indicating a pro-proliferative response. Esr1 and Gata3 expression were also clearly induced by post OVX IF exposure, alone and in combination with additional E2 exposure. Esr1 upregulation might explain the increased estrogenic response. However, a recent animal study did not show pro-estrogenic and proliferative effects on the MGs of rats due to a post OVX IF exposure (150 mg/kg bw) as determined by IHC of Ki67 [265].

In comparison to lifelong IF exposure, post OVX IF intervention significantly affected DNA methylation in the opposite direction. This was the case for Cldn4, Extl1, Gsn, Niacr1, PTEN and Sirt4 which were hypomethylated in the CH diet group and not affected or hypermethylated in the HH group. Further investigations have to address mechanistic aspects underlying this differential epigenetic response. Consistent with methylation changes in the promoter region, Cldn4 mRNA expression was induced in the CH group and downregulated in the HH group. High Cldn4 expression was recently detected in triple negative breast cancer, a breast cancer subtype with poor prognosis. Therefore, increased Cldn4 expression due to the post OVX IF exposure can be considered as an adverse effect regarding MG health [266]. Consistently, expression of the TSG PTEN was significantly reduced in the CH group, but not in the HH group.

Taken together, our observations indicate a rather pro-proliferative, pro-estrogenic and pro-carcinogenic role of a post OVX exposure to high IF levels compared to the more preventive effects of a lifelong exposure. Although these results cannot be directly translated to the human situation, uptake of high doses of IF as supplements after menopause should be considered with some caution [267-269].

5.2.2 Investigations on the kinetics of DNA methylation and on carcinogenesis in ACI rats

In the second part of our investigations, we focused on the impact of a lifelong dietary IF exposure on the kinetics of DNA methylation changes in MGs of ACI rats during all developmentally critical time windows, *e.g.*, embryonic/fetal, early postnatal, and pubertal periods. MG samples were analyzed and information on UWW was obtained from animals at PND 21, 50, 80, 97 and 180. ACI rats are susceptible to E2-induced carcinogenesis. Therefore, we additionally focused on DNA methylation changes during MG tumorigenesis in ACI rats and the impact of lifelong IF exposure.

5.2.2.1 Kinetic experiment

Endocrinological parameters are not affected by lifelong high IF exposure

Lifelong high IF exposure in ACI rats resulted in a weak acceleration of vaginal opening, although this effect was not significant due to high interindividual variation especially in the control group. As described before, we and others have observed estrogenic effects of a prepubertal and multigenerational IF exposure indicated by a clearly accelerated onset of puberty in Wistar (1.5 days, first study of this thesis; 3 days, [235]), as well as in Sprague Dawley rats (3 days, [236]), and mice (6 days, [237]). Notably, vaginal opening in ACI rats was observed relatively late at a median age of about 46 days for the control group, in comparison to Wistar rat with vaginal opening at a median age of 38 days.

Relative UWW was affected by age, but not by dietary IF. Since UWW is depicted relative to bodyweight, this age-dependent effect rather reflects a gain in bodyweight with comparably slower uterine growth than an actual decrease in the absolute UWW.

Proliferative and estrogenic response is increased by lifelong high IF exposure in young rats

During normal development, proliferation of the MG increased as measured by Ki67 mRNA expression resembling the normal expansion of the ductal tree. Expression of Pgr as the prototypical marker for E2 responsiveness gradually increased with normal MG development as determined by mRNA expression as well as by IHC. Interestingly, IHC but not mRNA expression revealed an estrogenic effect of lifelong IF intake, identified as increased Pgr and Ki67 expression in the pre- and postpubertal ACI rats. These E2 agonistic effects could not be identified with mRNA analysis of these markers, however PCNA levels were increased at the prepubertal phase after rats were exposed to IF suggesting an acceleration of the MG development and a potential mechanism that might explain an altered susceptibility to tumor formation and probably the IF-reduced tumor incidence (discussed in detail in section 5.2.2.2).

The TEBs represent a critical target for tumor initiation, since they are highly undifferentiated and possess high proliferative capacity during puberty. Animals with prepubertal exposure to dietary IF had fewer TEBs on PND 50 when compared to the placebo group (Möller et al., 2016 in revision, [263]). This observation probably reflects an earlier peak of maximum TEB formation rather than reduced TEB formation [261]. Shortening the window of vulnerability by an early differentiation of the TEBs in response to IF exposure probably represents an important contribution to the overall mechanism by which IF exert their protective effects on breast cancer development. This is in line with Fritz et al. who showed that early exposure (gestational day 0–PND 21) to physiological doses of GEN (0, 25, 250 mg/kg diet) induced cell differentiation in Sprague Dawley rats [270].

In our study, ACI rats were all intact (not OVX) and not estrus cycle synchronized. Therefore, we had estrus, met-estrus, di-estrus, and pro-estrus phases, as well as transitions between the phases present, and plasma levels of endogenous E2 and its metabolites varied essentially. Normal MG epithelium undergoes oscillating phases of cell proliferation and cell death during the estrus cycle leading to phases of normal hyperplasia. Therefore, individual rats might have passed different parts of the estrous cycle which might contribute to interindividual differences in expression of proliferation markers and hormone receptors within one diet group.

Kinetics of DNA methylation is stronger affected by age than by diet

Epigenetic modifications accompany developmental processes and persist until later in life. Therefore, we investigated the kinetics of interesting changes of DNA methylation identified by RRBS with age groups PND 21 and 50 (discussed in section 5.1.2.1) from early development until adolescence and adulthood in ACI rats using MassARRAY technology.

MassARRAY analysis revealed significant age-dependent effects on DNA methylation with most pronounced effects during puberty. Methylation in the promoter regions of analyzed genes gradually changed up to 30 % starting from puberty onwards where lifelong IF exposure was not able to induce methylation changes within one age group. Methylation for *Niacr1* and *Fxyd2* was decreased while *Arpp21*, *Mmaa*, *Zfp507*, *Shroom2*, *Extl1* and *Esr1* underwent progressive hypermethylation. Interestingly, strongest effects were observed between pre- and postpubertal rats (PND 21 and 50) when hormone-dependent MG development is initiated.

Niacr1, is a G protein-coupled receptor located on adipocytes mediating anti-lipolytic effects and *Fxyd2*, as the gamma subunit of Na, K-ATPase are important regulators of whole body energy homeostasis [262] and cellular homeostasis [271]. Both are involved in adipose tissue metabolism and stem cell differentiation, respectively. It is well known that increased bodyweight and obesity are strongly associated with postmenopausal breast cancer [272]. Hypomethylation of these genes by IF

exposure might be able to affect the adipose tissue metabolism leading to beneficial effects by inducing the anti-lipolytic machinery. Arpp21 is frequently downregulated in breast cancer, where it is negatively associated with clinico-pathologic characteristics and survival outcome and might be an interesting target for IF exposure [247]. Since Arpp21 promoter methylation drastically increased after puberty it is tempting to speculate that E2-induced hypermethylation might be the underlying trigger for Arpp21 silencing in breast cancer cells.

Mmaa (methylmalonic aciduria (cobalamin deficiency) cblA type) is a protein causally involved in methylmalonic acidemia (MMA), an inborn error in vitamin B12 metabolism. Shroom2 proteins are essential regulators of cell shape and tissue morphology during animal development [273] and Zfp507 is a not further characterized zinc-finger protein that may be involved in transcriptional regulation. Although these genes underwent clear promoter hypermethylation with age, none of them has been implicated with human MG development or associated with breast carcinogenesis.

However, age-induced methylation changes did not affect the expression of any of the genes analyzed in the kinetic experiment.

Expression of epigenetic enzymes is modulated by age and IF exposure

In order to provide an underlying mechanism of the gradual age-induced changes in DNA methylation during MG development, we analyzed relative mRNA expression of enzymes modifying DNA methylation. The transcript levels of all three DNMTs showed tendencies of modification with increasing age. DNMT1 expression was gradually increased whereas expression of DNMT3a and DNMT3b was lower at older age. This is in line with a study from Kovalchuk et al. who investigated ACI rats early after a subcutaneous E2 treatment in comparison to normal MG [212]. DNMT3a levels measured by IHC were gradually downregulated with increasing age in the healthy MG (PND 47-75), which in turn might mediate hypomethylation and enables binding of additional DNA- or chromatin-modifying enzymes that are sensitive to DNA methylation. Accordingly, Kutanzi et al. verified a concomitant increase of H4K12Ac in MGs of aging ACI rats. H4K12Ac is a transcriptional active histone mark, involved in histone disposition and was shown to occupy E2-responsive gene promoters and EREs in ER α -positive MCF-7 cells in an inducible manner [274].

Lifelong IF intervention reversed the observed age-related alterations towards a bell-shaped pattern for DNMT1 and lower expression levels for DNMT3a and DNMT3b at PND 21 in the H compared to the C group. Since we did not see IF-induced effects on DNA methylation at PND 21 or 50, expression of the DNA methylation modifying enzymes might be of minor impact on epigenetic regulation of the investigated candidate regions. However, Tet1 mRNA expression was clearly reduced in the course of

MG development in both dietary groups. A reduced Tet1 activity probably leads to less active demethylation and lower 5hmC levels, and is linked to gene body hypermethylation a phenomenon that has been observed for several cancer types including breast cancer [66, 275]. Since our analyses are based on bisulfite converted DNA which does not distinguish 5hmC from 5mC, we are not able to detect this loss of 5hmC. Further manipulation such as chemical oxidation of 5hmC to 5fC which then reacts with sodium bisulfite and is identified as unmethylated C would be necessary to allow a precise readout of 5mC by bisulfite sequencing (oxBS-Seq). 5hmC levels can then be deduced by comparing oxBS-Seq (which identifies 5mC only) with BS-Seq (which identifies both, 5mC and 5hmC) [77]. Clearly, investigation of 5hmC levels would be interesting, since this DNA modification has been proposed to contribute to the transcriptional regulation of active and bivalent gene promoters [276].

DNA methylation changes reflect alterations in glandular structure

Normal MG development involves a massive reorganization of the MGs cytoarchitecture. EGF controls early duct outgrowth and together with E2 promotes massive remodeling of the breast architecture. Ducts grow, divide and form TEBs which are highly proliferative structures that give rise to branches, ductules, and to a lesser extent alveolar buds with enhanced degree of differentiation. With increasing age, ductal elongation and infiltration of the surrounding fat stroma also increase the percentage of the ducts in the MGs [108]. Therefore, methylation differences of candidate regions induced by age probably just reflect the changes in tissue composition due to MG development.

We showed that age-dependent epigenetic effects were observed mainly between PND 21 and 50. These changes were induced by endogenous E2 exposure. Ding et al. recently demonstrated that exposure of ACI rats to exogenous E2 for 1 week induced profound changes in proliferation and cytoarchitecture [194]. Therefore, we characterized the cellular composition of MGs during development and determined gene expression patterns for selected marker genes for epithelial cells (Krt8 and Cdh1) as well as myoepithelial cells (Krt5, SMA and Tp63). Analyzing mRNA expression of the epithelial cell markers, we observed increased transcript levels from puberty onwards with pronounced effects during the pubertal phase, indicating the proliferative activity during MG development. This observation fits to the gradually increasing gene and protein expression levels of the proliferation markers (PCNA and Ki67). Since ductal outgrowth and branching is vastly driven by the hormonal interplay of E2 and progesterone, epithelial cells expressed Pgr, Esr1 and Esr2 with only Pgr being highly upregulated from puberty onwards. This has been reviewed by Hennighausen et al. who described an increased Pgr expression of proliferative cells and that a part of E2-mediated effects were due to Pgr induction and did not depend on Esr1 or Esr2 upregulation [277].

Therefore, the gradually modified methylation patterns might be, indeed, a result of the increasing numbers of epithelial cells, contributing to the analyzed specimen rather than an actual change in methylation within the epithelial cells.

Observation from gene expression studies could support this hypothesis since none of the genes investigated with significant promoter DMRs also resulted in gene expression alteration. mRNA level are normalized to housekeeping genes such as Hprt, TBP and β -actin which reflect gene expression patterns that arise due to normal development. Therefore, detection of tissue composition-related transcriptional changes might be lost, since relative mRNA expression is calculated. This would further explain the fact that changes in DNA methylation did not correlate with gene expression of the corresponding candidate regions.

For further analyses, purification of specific MG cell populations (*e.g.*, epithelial cells) prior to analysis of DNA methylation and gene expression by cell-surface expressing markers or specific digestion with collagenase/hyaluronidase would substantially increase the understanding of the diversity of epigenetic regulations in individual cell types. Additionally, it would be interesting to investigate the kinetics of epigenetic changes after exogenous E2-exposure to identify the earliest methylation changes as potential epigenetic drivers/triggers of carcinogenesis, since epigenetic instability has been described to precede genomic instability in the multistage process of breast carcinogenesis [278].

5.2.2.2 Carcinogenesis experiment

The ACI rat is an animal model susceptible to develop MG tumors upon prolonged exposure to pregnancy levels of exogenous E2 [279] and is a highly relevant model with a close connection to human types of breast cancer. Distinct quantitative trait loci, *i.e.*, estrogen-induced mammary cancer (Emca) underlie the breast cancer susceptibility with several orthologous loci in human [166] being associated with risk factors such as mammographic breast density [167].

It has been hypothesized by Barker et al. that early nutritional exposures are associated with disease onset later in life [280]. His theory of a fetal origin of disease has now widely expanded to the hypothesis of 'Developmental Origins of Health and Disease' (DOHaD) [281], linking an adverse developmental environment with increased risk of disease in adulthood. To date hormone dependent malformation of the reproductive tract of male and female have provided the strongest evidence for a developmental reprogramming and alteration of cancer susceptibility induce by, *e.g.*, Diethylstilbestrol and to a less severe extent Bisphenol A [282]. Also limb mal formation caused by exposure to Thalidomide or higher incidences of coronary heart disease and obesity after famine exposure during the first trimester are well known examples [282]. Phytoestrogenic soy IF have been

shown to modulate MG morphology and provide anti-tumorigenic effects especially when exposed in a multigenerational setting [141]. Since *in utero* and additionally prepubertal exposure link fetal exposures and epigenetic programming, we assessed the cancer preventive impact of a lifelong dietary IF exposure on E2-driven carcinogenesis of the MG.

IF exposure reduced incidence of E2-induced MG tumors but accelerated onset and growth

Lifelong IF exposure showed a dualistic mode of action since it reduced incidence and multiplicity of MG tumors but shortened tumor latency and enhanced tumor growth if tumors escaped the preventive effect. This can be interpreted as an inhibitory impact of IF on the tumor initiation process, but also as a tumor promoting activity of IF on the E2-induced tumors.

Since final tumor incidence was slightly reduced it can be stated that dietary IF intake protected against E2-induced breast carcinogenesis although tumor growth was not prevented. This is in line with Lamartiniere et al. investigating DMBA-induced MG tumors in female Sprague Dawley rats. It was shown that a prepubertal GEN exposure (250mg genistein/kg diet) reduced the number of DMBA-induced MG tumors through a decline in the number of TEBs at PND 50 of female rats associated with a decreased rate of cell proliferation by more than 50 % [283]. However, this protective action of GEN against chemically induced tumors was not detectable if administered prenatally only [263], highlighting that there are sensitive phases for beneficial dietary interventions which might be more represented by the postnatal MG development stages.

However, the data also suggest that chronic intake of IF reduced the time-to-tumor appearance and promoted growth of initiated tumors. Similar findings by Watson et al. support these results. Here, MTB-IGFIR transgenic mice that develop MG tumors upon doxycycline-induced overexpression of the IGF receptor were chronically exposed to IF from conception onwards. Tumor incidence and onset was significantly increased in the IF group. Since markers of E2 signaling (*e.g.*, Pgr, Areg) were upregulated, the authors assumed an ER-signaling dependent mode of action [284]. When comparing these finding with xenograft models, IF intervention was shown to induce growth of implanted ER α -positive MCF-7 cells but not ER α -negative MDA-MB-231 cells in ovariectomized athymic mice [285, 286]. Summarizing existing data suggest that dietary IF might stimulate growth of preexisting MG tumors in rodents but these adverse effects might be limited to ER α -dependent tumor cells.

Estrogenic response is not affected by lifelong high IF exposure in adult rats

During normal development proliferation of the MG increased as measured by Ki67 mRNA expression, resembling the normal expansion of the ductal tree. Further increase during tumorigenesis represents aberrant growth of the MG epithelium and expansion of the cell volume, an observation that was confirmed by IHC in MGs of E2-exposed animals that did not develop

palpable tumors. In these glands, E2 stimulates the expression of the proliferation marker Ki67 before the development of pre-neoplastic lesions. Similar enhanced cell proliferation in mild to moderate hyperplastic MGs was observed by Kutanzi et al. who investigated E2-treated ACI rats at PND 100, 140 and 180 [287]. Additionally, the authors demonstrated an imbalance between cell proliferation and apoptosis in MG/T at PND 100 rather than an enhanced proliferation or suppressed apoptosis alone as the driving force towards neoplasia. Further, Mdm2, a proto-oncogene was sustainably upregulated from mild hyperplasia onwards, together with a significant increase in c-Myc expression. Taken together the data suggest that the dysbalance of cellular processes that trigger uncontrolled cell proliferation, genomic instability, and neoplastic transformation is detectable at early stages in E2-treated MGs and independent of the prevalence of tumors.

Pgr as a marker for E2 responsiveness gradually increased with normal MG development and got massively upregulated by E2 treatment as determined by mRNA expression as well as by IHC. However, transcript levels of both Esr1 and Esr2 were rather downregulated in tumors independent of the diet administered. These findings support the idea that the increase in Pgr expression is accompanied with proliferative cells and is responsible for the E2-mediated effects that are, indeed, not derived from Esr1 and Esr2 upregulation [277].

The increase in transcripts and protein levels of the proliferation markers Ki67 and PCNA in tumor tissue could not explain the growth promoting effects of IF in a subset of tumors since these effects were independent of the dietary regime. However, single tumors of the H group showed a massive gain in Ki67 mRNA expression probably resembling those with highest tumor weights.

DNA methylation and gene expression of candidate genes

We performed RRBS to identify DMS related to either dietary exposure to IF, to carcinogenesis, or a combination of both (discussed in section 5.1.2.2). We selected eleven candidate genes for validation by MassARRAY. For six out of 11 genes, methylation levels correlated with expression levels. Most DMRs were annotated to promoter regions, emphasizing the relationship between promoter methylation and regulation of gene transcription.

Methylation data for St3gal4 (ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (sial4c), a gene involved in sialylation, which plays a crucial role in cell adhesion and regulates the biological stability of glycoproteins [193] was determined for an intergenic DMR roughly 138.4 kb downstream of the TSS with significantly correlating transcript levels. This DMR might overlap with an enhancer region which regulates gene expression of the target gene. So far, different from the human and mouse genomes, the exact mapping of enhancers in the rat genome is not well known.

Although methylation levels did not significantly change, mRNA transcript levels of Zfp507, a not further characterized zinc-finger protein, were low in placebo group tumors and expression was further reduced in tumors of the H diet group. Additionally, Mrc2 (receptor, C type 2), involved in the collagen turnover process, Thy (Thy-1 cell surface antigen), affecting cell-cell or cell-matrix interactions as well as Rnf8 (ring finger protein 8), an E3 ubiquitin protein ligase involved in chromatin remodeling underwent significant carcinogenesis related methylation and gene expression changes. However, we did not detect significant correlations in the carcinogenesis experiment indicating that additional epigenetic mechanisms such as histone tail modifications and non coding RNAs might account for these expression changes.

However, mRNA levels of Fxyd2 downregulated during adipose stem cell differentiation and involved in basic cellular homeostasis, and Arpp21 increased in MG of the H diet-exposed animals vs. the placebo group, and were significantly repressed in tumors of both dietary groups. Since Arpp21 has been reported to be downregulated in 87 % of primary breast cancers, IF-induced upregulation would be a clear beneficial effect. No IF impact was detectable for relative Esr1 and Niacr1, a receptor located on adipocytes and involved in anti-lipolytic effects, but levels were significantly decreased due to the E2 treatment, supporting the idea of a tissue-specific autoregulatory mechanism of the Esr1 expression upon exogenous E2 stimulation [252].

Due to the breast carcinogenic process mRNA levels of epithelial cell markers Krt8 and Cdh1 got massively upregulated while myoepithelial markers Krt5, SMA and TP63 were downregulated. This opposing expression represents the E2-stimulated clonal expansion of luminal cells in the MG during the breast carcinogenic process. Similarly to the observed effects in the pubertal ACI rat study, methylation and gene expression differences of the candidate regions might be attributed to changes in tissue composition due to tumor development and vastly reflect the actual methylation changes in the hyperplastic epithelial cells. However, characterizing the average methylation across the bulk of cell populations certainly conceals cell type specific methylation profiles. The identified small changes in gene expression or methylation might originate from, *e.g.*, a myoepithelial cell fraction which is basically underrepresented in this approach and, thereby, cell type specific effects are "diluted". Gascard et al. investigated different enriched primary human breast cell types from disease free women and defined comprehensive differences in epigenomic reprogramming between luminal and myoepithelial cells such as methylation of enhancer elements and transcriptional output (RNA yield/cell) [288]. Therefore, purification of specific MG cell populations based on cell-surface expressing markers or possibly even single-cell epigenomics could substantially unravel the diversity in the individual cellular states at its native resolution [289].

Writers and erasers of DNA methylation

In order to identify the underlying mechanism of the DNA methylation and gene expression changes during tumor development, we investigated effects on epigenetic modifiers induced by the environmental stimuli and endocrine signals. We selected epigenetic modifying enzymes involved in writing and erasing DNA methylation, *i.e.*, DNMT1, 3a, and 3b, as well as Tet1.

The relative mRNA expression of DNA methylation modifying enzymes show a diverse pattern. During early MG development DNMT3a and 3b were significantly downregulated by lifelong IF intervention and were further decreased during tumor formation. DNMT1 expression was gradually increased during normal development with a bell-shaped pattern in the IF group and got further upregulated during tumor formation. This phenomenon was also observed by others, where higher levels were also present in the pre-neoplastic lesions of E2-treated ACI rats [168, 212]. However, lifelong IF exposure prevented the E2-induced upregulation of DNMT1 in tumors. This is in line with several observations regarding the potential of IF to alter DNMT activity by, *e.g.*, interacting with the catalytic domain and subsequent suppression of the DNMT activity [156].

Transcript levels of Tet1 were also clearly reduced in the course of MG development and E2-induced tumor formation, an effect which was most pronounced in MG/T. This is in line with observations in colon cancer where Tet1 downregulation was detectable in very early stages of cell transformation [290]. Interestingly, mRNA levels of Tet1 and DNMT3b as well as DNMT1 are clearly opposing during MG development and tumor formation, respectively and most pronounced due to the IF exposure, demonstrating their roles as counterparts in shaping the DNA methylation landscape.

The IF-mediated downregulation of DNMTs possibly inhibits the process of regional hypermethylation of promoter and enhancer regions of putative TSGs during tumor formation. The consequence of this prevention might be an increased accessibility of the transcription machinery and increased TSG protein levels. Therefore, IF-reduced prepubertal DNMT activity (as shown for DNMT3a and 3b mRNA levels) might alter the susceptibility of the MG towards tumorigenesis by reprogramming the pubertal methylome.

Functional pathways affected by lifelong IF exposure

Tissue development, organogenesis and cellular differentiation occur through a series of tightly regulated, temporally and spatially timed molecular and biochemical events [282]. Lifelong IF exposure reduced the incidence and multiplicity of MG tumors and induced more than 24,000 DMS between the tumors of E2-treated ACI rats (comparison C+E2 vs. H+E2). Using Ingenuity Pathway Analysis (IPA) we observed intriguing enrichments of DMS in distinct biological processes. Environmental IF intervention during development affected genes of several functional pathways

and reprogrammed tissue responses, *e.g.*, in reproductive development and function, reproductive system disease, as well as in endocrine system development and function. Further, genes were enriched in cellular events such as growth and proliferation, development, movement, signaling, and interaction, as well as morphology *etc.*

We identified differentially methylated genes that are related to several of these signaling pathways and that might be interesting targets for further analyses. Hypermethylation was identified for DMRs associated with: RAX, retina and anterior neural fold homeobox, a homeodomain-containing TF with functions in eye development which has not been implicated in tumor development, yet [291]. SMAD7, a nuclear protein that binds to E3 ubiquitin ligase SMURF2 and has been shown to promote EGFR signaling, *via* impairing ligand-induced ubiquitination and degradation of activated EGFR [292, 293]. MafK that belongs to bZIP-domain containing TFs might be involved in regulation of NF- κ B, of NF-E2 (Nrf2) which binds to the antioxidant-responsive element (ARE), and in production of ROS [294], [295]. Interestingly, MafK popped up in the TF enrichment analysis in the Wistar rats, as well. Therefore, the mechanistic impact of the bZIP-domain containing TFs might indeed represent a promising target for IF-mediated transcriptional regulation. Additionally Wnt5a, a ligand of the non-canonical Wnt-pathway was highly methylated. This is in line with Su et al. who were interested in the potential biological and molecular pathways altered by IF in isolated mammary epithelial cells (MEC) of Sprague Dawley rats (PND 50). Using microarray technology they identified various biochemical pathways affected after administration of GEN- and SPI-enriched diets (250 and 400 ppm IF aglycone equivalents, respectively). They provide evidence for the downregulation of the proto-oncogene Wnt5a [296] as well as Zhang et al. who identified a decreased AOM-induced Wnt5a expression in the colon due to a SPI (140 ppm GEN) intervention in Sprague-Dawley rats [297]. The Wnt-pathway is well known for its implications in colorectal cancer however the current results suggest a pivotal role in IF-mediated effects in MG carcinogenesis.

Hypomethylated DMRs have been observed for: Trip11, thyroid hormone receptor interactor 11 which has not yet been implicated in carcinogenesis and HOXD3, a homeodomain-containing transcription factor of the HOX family that determines cellular identity during development. Hypermethylation of HOXD3 is associated with clinico-pathologic features such as progression in prostate cancer and shorter survival in invasive breast cancer [298], [299]. Therefore, a demethylation of the HOXD3 promoter might be of beneficial effect regarding MG tumor development. Further, the promoter of PIK3R1, phosphoinositide-3-kinase regulatory subunit 1 harbored a hypomethylated DMR. It has been reported that IF partially exert their cancer preventive properties by targeting the PI3K/AKT signaling pathway. Bredfeldt et al. identified an interesting and rapid non-genomic signaling pathway upon binding of E2 or the endocrine disruptor Diethylstilbestrol

which however, is not fully understood, yet [300]. Ligand binding to the membrane-associated ER activates PI3K/Akt signaling and phosphorylates the histone methyltransferase enhancer of Zeste homolog 2 (EZH2) which adds methyl groups to the lysine 27 residue of histone 3 tails. The phosphorylation leads to an inactivation of EZH2 and a concomitant reduction of H3K27me3 levels, a transcriptional repressive histone mark in ER α -positive MCF-7 cells, uterine myometrial cells (ELT3) and in the postnatal myometrium of Eker rats. However, in a recent publication of Greathouse et al. subcutaneous injections of GEN (50 mg/kg per rat) from PND 10-12 induced the same effects in the uterus of Eker rats [301]. Furthermore, GEN developmentally reprogrammed E2-responsive genes to become hyper-responsive. Although DNA methylation levels were not particularly assessed in these studies the observed hypomethylation of the PIK3R1 promoter and a possible subsequent upregulation demonstrate how IF can regulate epigenetic mechanisms *via* DNA methylation changes.

Since IF caused developmental reprogramming and affected control variables in many pathways it is tempting to speculate that several of these signals are responsible for the identified tumor preventive efficacy observed for IF containing diet. However, our results also identified a tumor promoting activity of IF on the E2-induced tumors that escaped the preventive effect. Since we compared frank tumors of either group for our analysis, primarily the physiological signals responsible for the adverse effects might be encompassed in the identified pathways.

5.2.2.3 Quantitative DNA methylation analyses of enriched MG epithelial cells

Validating the genome-wide data with MassARRAY technology, we observed that the cellular composition of the MGs might contribute to differences observed in DNA methylation measurements and complicates the identification of diet-induced effects. Therefore, it was desirable to assess homogenous epithelial cells. Since isolation of epithelial cells by flow cytometry using defined surface markers was not an option due to the nature of the specimen (cryo-sections), we performed laser capture microdissection to enrich for cells originating from the ductal tree.

In contrast to our expectations, interindividual DNA methylation was often greater in the LCM samples when compared to samples obtained from the MG bulk of cells. Within one diet group methylation values showed massive divergence which did not enable a clear insight into epithelial DNA methylation changes unimpacted by other cell subtypes.

As mentioned previously, bisulfite conversion commonly destroys >99 % of intact DNA fragments due to the denaturing environment. Since the amount of input material was already very low (<30 ng) only a few fragments were available for subsequent amplifications during MassARRAY PCR. Since this protocol involved more than 40 amplification cycles the residual intact fragments are amplified but do not necessarily represent the initially captured entirety of fragments. High duplication rates of un-

or fully methylated DNA fragments have been reported leading to a tendency to observe either 0 %, 50 % or 100 % methylation (oral communication PD Dr. Dieter Weichenhan, Christoph Weigel). Thus, the technical artifacts might introduce less bias when fragments are analyzed by a different quantitative technique and PCR induced duplicates, amplicons with the exact same sequence pattern of CpG and non-CpG cytosines, are filtered out prior to the DNA methylation analysis, *e.g.*, .bisulfite sequencing.

6. Conclusions and outlook

In the present thesis, two individual sequencing-based methodologies for methylation profiling, *i.e.*, Methyl-CpG immunoprecipitation-sequencing (MCIp-Seq) and reduced representation bisulfite sequencing (RRBS) were implemented to achieve comprehensive insights into the genome-wide effects of IF on DNA methylation in the rat MG. Animals experiments were performed using three IF-enriched diets containing low, medium and high concentration of IF, based on the reported dietary IF intake of the Western and Asian population.

We identified that i) IF are able to modify DNA methylation patterns in the MGs of rats on a genome-wide scale, although with smaller absolute methylation changes compared to those observed during, *e.g.*, carcinogenesis, ii) the impact of IF on DNA methylation was highly dependent on the IF dose administered as well as on iii) the timing of exposure, and iv) the chemopreventive efficacy on E2-induced breast carcinogenesis indicated a dualistic mode of action.

Using MCIp-Seq and RRBS, we identified IF-mediated methylation changes, including both hypo- and hypermethylated regions. More than 90 % of the differential methylation occurred outside of classical regulatory regions, but also included differential methylation at specific TF binding sites. Therefore, we assume that IF predominantly exert their functional roles by engaging distal regulatory elements and by modifying TF regulated gene transcription. So far, there is not much known about enhancers and distal promoters in the rat genome. Therefore, it would be of great interest to investigate specific histone tail modifications such as H3K27ac, H3K4me1 and H3K4me3 in order to identify genomic locations of active distal enhancers and/or promoters in the rat MG, for an integrative analysis in combination with our methylation data. Differentially methylated promoters have been associated with signaling pathways in reproductive tissue homeostasis and endocrine system. For future studies, transcriptomic data for comparative analysis would further improve our understanding of the biological impact of a dietary IF exposure on epigenetic regulation.

Comparing different lifelong IF exposure protocols we observed a biphasic dose response profile, with the low dietary IF concentration being not sufficient to reduce estrogenic and proliferative response (measured by PR and PCNA protein expression). Overall, IF induced U-shaped dose-dependent DNA methylation and mRNA expression changes. These results allow the conclusion that high IF levels are needed to achieve potential beneficial health effects. High IF intake (>100 mg IF/day) is a phenomenon observed for approximately 10 % of the Asian population, represented by the diets containing higher IF concentrations in our study (corresponding to a daily intake of about 177 mg and 390 mg IF, respectively, calculated for a 60 kg woman). Therefore, the reduced breast

cancer risk as seen for the Asian population might only be in parts be explained by the anti-proliferative and anti-estrogenic effects of a lifelong exposure to high IF concentrations.

In contrast to lifelong exposure to the highest IF concentration, intervention exclusively during puberty showed no clear impact on proliferation markers and exhibited only minor effects on DNA methylation and mRNA expression levels. We assume that the pubertal period is not the most critical time window responsible for introducing long-term chemopreventive effects. Earlier developmental windows such as the postnatal or prepubertal phase might therefore hold the key for preventive epigenetic reprogramming of the MG and should be taken into consideration for further preventive approaches. However, a short term exposure to high IF levels after menopause, as often observed in Western countries as alternative to HRT, displayed pro-proliferative as well as pro-estrogenic properties in MGs of adult rats. Contraindicative effects were also identified for DNA methylation and mRNA expression patterns since they were affected in an opposed direction when compared to lifelong exposure to the highest IF concentration. Although our results do not permit a clear risk assessment regarding the commercially available natural alternatives to HRT, supplementation of high doses of IF after menopause should not be recommended without any concerns.

With respect to lifelong IF-mediated effects on E2-induced rat mammary carcinogenesis, we identified a dualistic mode of action. Incidence and multiplicity of MG tumors were reduced, but we observed a shortened tumor latency and enhanced tumor growth if tumors escaped the preventive effect. Further, lifelong exposure to high IF levels induced estrogenic and proliferative response of the MGs at puberty, as measured by PR and Ki67 protein expression. We assume that the pro-proliferative effects during early MG development, in contrast to later stages, might be interpreted as an advantage regarding MG health. Since an acceleration of the MG differentiation shortens the window of vulnerability of the MG, these early pro-estrogenic effects might provide a mechanism by which IF exert their chemopreventive effects and reduce the susceptibility of the MGs towards the subsequent E2 treatment. Epigenetic modifying enzymes, such as DNMT3a and 3b were downregulated by lifelong IF exposure from early MG development onwards and represent interesting targets for IF to shape the DNA methylation landscape. Since a number of the promoter DNA methylation changes that we investigated did not result in gene expression alterations of the target genes it would be of great interest to investigate the corresponding chromatin-modifying enzymes as well as putative non-coding RNAs such as miRNAs using, for example, nCounter Nanostring technology or RNA-Seq. This knowledge would enable a comprehensive overview of the epigenetic modifying potential of a lifelong soy IF exposure on MG health and disease.

Our results cannot be directly translated to the human situation since rodent IF, and E2 metabolism, as well as microbial intermediates differ substantially. The human relevance of our data needs to be clarified in further studies; however, with our results in hands, we do not unreservedly recommend the use of IF-containing capsules as a safe alternative to HRT.

7. References

1. Nutrient Data Laboratory, Beltsville Human Nutrition Research Center, and U.S.D.o.A. Agricultural Research Service. *USDA Database for the Isoflavone Content of Selected Foods, Release 2.0* 2008 [cited 2015; Available from: http://www.ars.usda.gov/SP2UserFiles/Place/80400525/Data/isoflav/Isoflav_R2.pdf.
2. Pudenz, M., K. Roth, and C. Gerhauser, *Impact of Soy Isoflavones on the Epigenome in Cancer Prevention*. *Nutrients*, 2014. **6**(10): p. 4218-4272.
3. Chan, S.G., et al., *Dietary sources and determinants of soy isoflavone intake among midlife Chinese Women in Hong Kong*. *The Journal of Nutrition*, 2007. **137**(11): p. 2451-5.
4. Seow, A., et al., *Isoflavonoid levels in spot urine are associated with frequency of dietary soy intake in a population-based sample of middle-aged and older Chinese in Singapore*. *Cancer Epidemiology Biomarkers and Prevention*, 1998. **7**(2): p. 135-140.
5. Keinan-Boker, L., et al., *Soy product consumption in 10 European countries: the European Prospective Investigation into Cancer and Nutrition (EPIC) study*. *Public Health Nutrition*, 2002. **5**(6b): p. 1217-1226.
6. Setchell, K.D. and E. Lydeking-Olsen, *Dietary phytoestrogens and their effect on bone: evidence from in vitro and in vivo, human observational, and dietary intervention studies*. *The American Journal of Clinical Nutrition*, 2003. **78**(3 Suppl): p. 593S-609S.
7. Nurmi, T., et al., *Isoflavone content of the soy based supplements*. *Journal of Pharmaceutical and Biomedical Analysis*, 2002. **28**(1): p. 1-11.
8. Bolca, S., et al., *Disposition of soy isoflavones in normal human breast tissue*. *The American Journal of Clinical Nutrition*, 2010. **91**(4): p. 976-84.
9. Franke, A.A., J.F. Lai, and B.M. Halm, *Absorption, distribution, metabolism, and excretion of isoflavonoids after soy intake*. *Archives of Biochemistry and Biophysics*, 2014. **559**: p. 24-8.
10. Mortensen, A., et al., *Analytical and compositional aspects of isoflavones in food and their biological effects*. *Molecular Nutrition and Food Research*, 2009. **53 Suppl 2**: p. S266-309.
11. Grosser, G., et al., *Transport of the soy isoflavone daidzein and its conjugative metabolites by the carriers SOAT, NTCP, OAT4, and OATP2B1*. *Archives of Toxicology*, 2014.
12. Messina, M., *A Brief Historical Overview of the Past Two Decades of Soy and Isoflavone Research*. *The Journal of Nutrition*, 2010. **140**(7): p. 1350S-1354S.
13. Spagnuolo, C., et al., *Genistein and Cancer: Current Status, Challenges, and Future Directions*. *Advances in Nutrition: An International Review Journal*, 2015. **6**(4): p. 408-419.
14. Taylor, C.K., et al., *The effect of genistein aglycone on cancer and cancer risk: a review of in vitro, preclinical, and clinical studies*. *Nutrition Reviews*, 2009. **67**(7): p. 398-415.
15. D'Adamo, C.R. and A. Sahin, *Soy foods and supplementation: a review of commonly perceived health benefits and risks*. *Alternative Therapies In Health And Medicine*, 2014. **20 Suppl 1**: p. 39-51.
16. Riyazi, S.R., et al., *Chemical properties, health benefits and threats of soyisoflavones*. *Annals of Biological Research*, 2011. **2**(5): p. 338-350.
17. Vitale, D.C., et al., *Isoflavones: estrogenic activity, biological effect and bioavailability*. *European journal of drug metabolism and pharmacokinetics*, 2013. **38**(1): p. 15-25.

18. Sirotkin, A.V. and A.H. Harrath, *Phytoestrogens and their effects*. European Journal of Pharmacology, 2014. **741**: p. 230-236.
19. Steiner, C., et al., *Isoflavones and the prevention of breast and prostate cancer: new perspectives opened by nutrigenomics*. British Journal of Nutrition, 2008. **99**(E-S1): p. ES78-ES108.
20. Hervouet, E., et al., *Epigenetic regulation of estrogen signaling in breast cancer*. Epigenetics, 2013. **8**(3): p. 237-45.
21. Thomas, C. and J.A. Gustafsson, *The different roles of ER subtypes in cancer biology and therapy*. Nature Reviews Cancer, 2011. **11**(8): p. 597-608.
22. Deroo, B.J. and K.S. Korach, *Estrogen receptors and human disease*. Journal of Clinical Investigation, 2006. **116**(3): p. 561-70.
23. Couse, J.F., et al., *Tissue distribution and quantitative analysis of estrogen receptor-alpha (ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic acid in the wild-type and ERalpha-knockout mouse*. Endocrinology, 1997. **138**(11): p. 4613-21.
24. Fang, H., et al., *Structure-Activity Relationships for a Large Diverse Set of Natural, Synthetic, and Environmental Estrogens*. Chemical Research in Toxicology, 2001. **14**(3): p. 280-294.
25. Kuiper, G.G., et al., *Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta*. Endocrinology, 1998. **139**(10): p. 4252-63.
26. Smith, C.L. and B.W. O'Malley, *Coregulator function: a key to understanding tissue specificity of selective receptor modulators*. Endocrine Reviews, 2004. **25**(1): p. 45-71.
27. Riggs, B.L. and L.C. Hartmann, *Selective estrogen-receptor modulators -- mechanisms of action and application to clinical practice*. The New England Journal of Medicine, 2003. **348**(7): p. 618-29.
28. Oseni, T., et al., *Selective estrogen receptor modulators and phytoestrogens*. Planta Medica, 2008. **74**(13): p. 1656-65.
29. D'Adamo, C.R. and A. Sahin, *Soy foods and supplementation: a review of commonly perceived health benefits and risks*. Alternative Therapies In Health And Medicine, 2014. **20**(1): p. 39-51.
30. Rossouw, J.E., et al., *Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial*. The Journal of the American Medical Association, 2002. **288**(3): p. 321-33.
31. Poluzzi, E., et al., *Phytoestrogens in postmenopause: the state of the art from a chemical, pharmacological and regulatory perspective*. Current Medicinal Chemistry, 2014. **21**(4): p. 417-36.
32. Howes, L.G., J.B. Howes, and D.C. Knight, *Isoflavone therapy for menopausal flushes: a systematic review and meta-analysis*. Maturitas, 2006. **55**(3): p. 203-11.
33. Thomas, A.J., et al., *Effects of isoflavones and amino acid therapies for hot flashes and co-occurring symptoms during the menopausal transition and early postmenopause: a systematic review*. Maturitas, 2014. **78**(4): p. 263-76.
34. Blum, S.C., et al., *Dietary soy protein maintains some indices of bone mineral density and bone formation in aged ovariectomized rats*. The Journal of Nutrition, 2003. **133**(5): p. 1244-9.
35. Yu, F., et al., *Soybean isoflavone treatment induces osteoblast differentiation and proliferation by regulating analysis of Wnt/beta-catenin pathway*. Gene, 2015. **573**(2): p. 273-7.

36. Cosman, F., et al., *Bone density change and biochemical indices of skeletal turnover*. Calcified Tissue International, 1996. **58**(4): p. 236-43.
37. Baglia, M.L., et al., *Soy isoflavone intake and bone mineral density in breast cancer survivors*. Cancer Causes and Control, 2015. **26**(4): p. 571-80.
38. Cassidy, A., *Dietary phytoestrogens and bone health*. The journal of the British Menopause Society, 2003. **9**(1): p. 17-21.
39. Ma, D.F., et al., *Soy isoflavone intake inhibits bone resorption and stimulates bone formation in menopausal women: meta-analysis of randomized controlled trials*. European Journal of Clinical Nutrition, 2008. **62**(2): p. 155-61.
40. Jenkins, D.J.A., et al., *Soy Protein Reduces Serum Cholesterol by Both Intrinsic and Food Displacement Mechanisms*. The Journal of Nutrition, 2010. **140**(12): p. 2302S-2311S.
41. Hooper, L., et al., *Flavonoids, flavonoid-rich foods, and cardiovascular risk: a meta-analysis of randomized controlled trials*. The American Journal of Clinical Nutrition, 2008. **88**(1): p. 38-50.
42. Melendez, G.C., et al., *Beneficial effects of soy supplementation on postmenopausal atherosclerosis are dependent on pretreatment stage of plaque progression*. Menopause, 2015. **22**(3): p. 289-96.
43. Wiseman, H., et al., *Isoflavone phytoestrogens consumed in soy decrease F(2)-isoprostane concentrations and increase resistance of low-density lipoprotein to oxidation in humans*. The American Journal of Clinical Nutrition, 2000. **72**(2): p. 395-400.
44. Tikkanen, M.J., et al., *Effect of soybean phytoestrogen intake on low density lipoprotein oxidation resistance*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(6): p. 3106-10.
45. Beavers, D.P., et al., *Exposure to isoflavone-containing soy products and endothelial function: a Bayesian meta-analysis of randomized controlled trials*. Nutrition, Metabolism and Cardiovascular Diseases, 2012. **22**(3): p. 182-91.
46. Si, H. and D. Liu, *Genistein, a soy phytoestrogen, upregulates the expression of human endothelial nitric oxide synthase and lowers blood pressure in spontaneously hypertensive rats*. The Journal of Nutrition, 2008. **138**(2): p. 297-304.
47. Choudhuri, S., *From Waddington's epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research*. Toxicology Mechanisms and Methods, 2011. **21**(4): p. 252-74.
48. Jones, P.A. and S.B. Baylin, *The Epigenomics of Cancer*. Cell, 2007. **128**(4): p. 683-692.
49. Shen, H. and P.W. Laird, *Interplay between the cancer genome and epigenome*. Cell, 2013. **153**(1): p. 38-55.
50. Holliday, R., *A new theory of carcinogenesis*. British Journal of Cancer, 1979. **40**(4): p. 513-22.
51. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts*. Nature, 1983. **301**(5895): p. 89-92.
52. Feinberg, A.P. and B. Vogelstein, *Hypomethylation of ras oncogenes in primary human cancers*. Biochemical and Biophysical Research Communications, 1983. **111**(1): p. 47-54.
53. Laird, P.W., *Principles and challenges of genomewide DNA methylation analysis*. Nature Reviews Genetics, 2010. **11**(3): p. 191-203.

54. Bell, A.C. and G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene*. Nature, 2000. **405**(6785): p. 482-5.
55. Hark, A.T., et al., *CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus*. Nature, 2000. **405**(6785): p. 486-9.
56. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer*. Carcinogenesis, 2010. **31**(1): p. 27-36.
57. Santos, F., et al., *Dynamic reprogramming of DNA methylation in the early mouse embryo*. Developmental Biology, 2002. **241**(1): p. 172-82.
58. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer*. Nature Reviews Genetics, 2002. **3**(6): p. 415-28.
59. Goll, M.G., et al., *Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2*. Science, 2006. **311**(5759): p. 395-8.
60. Subramaniam, D., et al., *DNA methyltransferases: a novel target for prevention and therapy*. Frontiers in Oncology, 2014. **4**: p. 80.
61. Herman, J.G. and S.B. Baylin, *Gene silencing in cancer in association with promoter hypermethylation*. The New England Journal of Medicine, 2003. **349**(21): p. 2042-54.
62. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine*. Science, 2011. **333**(6047): p. 1300-3.
63. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*. Science, 2011. **333**(6047): p. 1303-7.
64. Oswald, J., et al., *Active demethylation of the paternal genome in the mouse zygote*. Current Biology, 2000. **10**(8): p. 475-8.
65. Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation*. Nature, 2013. **502**(7472): p. 472-9.
66. Yang, H., et al., *Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation*. Oncogene, 2013. **32**(5): p. 663-9.
67. Quintas-Cardama, A., F.P. Santos, and G. Garcia-Manero, *Therapy with azanucleosides for myelodysplastic syndromes*. Nature Reviews Clinical Oncology, 2010. **7**(8): p. 433-44.
68. Krell, D., et al., *IDH mutations in tumorigenesis and their potential role as novel therapeutic targets*. Future Oncology, 2013. **9**(12): p. 1923-35.
69. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex*. Nature, 1998. **393**(6683): p. 386-9.
70. Pandey, S., et al., *A novel MeCP2 acetylation site regulates interaction with ATRX and HDAC1*. Genes Cancer, 2015. **6**(9-10): p. 408-421.
71. Harrison, A. and A. Parle-McDermott, *DNA methylation: a timeline of methods and applications*. Journal of Genetics, 2011. **2**: p. 74, 1-13.
72. Plongthongkum, N., D.H. Diep, and K. Zhang, *Advances in the profiling of DNA modifications: cytosine methylation and beyond*. Nature Reviews Genetics, 2014. **15**(10): p. 647-661.

73. Stirzaker, C., et al., *Mining cancer methylomes: prospects and challenges*. Trends in Genetics, 2014. **30**(2): p. 75-84.
74. Bock, C., et al., *Quantitative comparison of genome-wide DNA methylation mapping technologies*. Nature Biotechnology, 2010. **28**(10): p. 1106-14.
75. Gerhauser, C., K. Heilmann, and M. Pudenz, *Genome-Wide DNA Methylation Profiling in Dietary Intervention Studies: a User's Perspective*. Current Pharmacology Reports, 2015. **1**(1): p. 31-45.
76. Wang, R.Y., C.W. Gehrke, and M. Ehrlich, *Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues*. Nucleic Acids Research, 1980. **8**(20): p. 4777-90.
77. Booth, M.J., et al., *Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine*. Nature Protocols, 2013. **8**(10): p. 1841-51.
78. Wang, Q., et al., *Tagmentation-based whole-genome bisulfite sequencing*. Nature Protocols, 2013. **8**(10): p. 2022-2032.
79. Eid, J., et al., *Real-time DNA sequencing from single polymerase molecules*. Science, 2009. **323**(5910): p. 133-8.
80. Gebhard, C., et al., *Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia*. Cancer Research, 2006. **66**(12): p. 6118-28.
81. Gebhard, C., et al., *Rapid and sensitive detection of CpG-methylation using methyl-binding (MB)-PCR*. Nucleic Acids Research, 2006. **34**(11): p. e82.
82. Klose, R.J., et al., *DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG*. Molecular Cell, 2005. **19**(5): p. 667-78.
83. Faryna, M., et al., *Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis*. The FASEB Journal, 2012. **26**(12): p. 4937-50.
84. Robinson, M.D., et al., *Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation*. Genome Research, 2010. **20**(12): p. 1719-29.
85. Nair, S.S., et al., *Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias*. Epigenetics, 2011. **6**(1): p. 34-44.
86. Meissner, A., et al., *Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis*. Nucleic Acids Research, 2005. **33**(18): p. 5868-77.
87. Gu, H., et al., *Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling*. Nature Protocols, 2011. **6**(4): p. 468-81.
88. Wang, J., et al., *Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing*. BMC Genomics, 2013. **14**: p. 11.
89. Boyle, P., et al., *Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling*. Genome Biology, 2012. **13**(10): p. R92.
90. Gu, H., et al., *Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution*. Nature Methods, 2010. **7**(2): p. 133-6.

91. Schillebeeckx, M., et al., *Laser capture microdissection-reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA methylation associated with gonadectomy-induced adrenocortical neoplasia in the mouse*. *Nucleic Acids Research*, 2013. **41**(11): p. e116.
92. Guo, H., et al., *Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing*. *Nature Protocols* 2015. **10**(5): p. 645-659.
93. Ehrich, M., et al., *Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry*. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. **102**(44): p. 15785-15790.
94. Tost, J., et al., *Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry*. *Nucleic Acids Research*, 2003. **31**(9): p. e50.
95. Kouzarides, T., *Chromatin modifications and their function*. *Cell*, 2007. **128**(4): p. 693-705.
96. Upadhyay, A.K. and X. Cheng, *Dynamics of histone lysine methylation: structures of methyl writers and erasers*. *Progress in Drug Research*, 2011. **67**: p. 107-24.
97. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications*. *Cell Research*, 2011. **21**(3): p. 381-95.
98. Kleer, C.G., et al., *EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells*. *Proceedings of the National Academy of Sciences of the United States of America*, 2003. **100**(20): p. 11606-11.
99. Metzger, E., et al., *LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription*. *Nature*, 2005. **437**(7057): p. 436-9.
100. Yang, S., et al., *KDM1A triggers androgen-induced miRNA transcription via H3K4me2 demethylation and DNA oxidation*. *The Prostate*, 2015. **75**(9): p. 936-46.
101. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis*. *Nature*, 2010. **464**(7291): p. 1071-6.
102. Prensner, J.R., et al., *The long noncoding RNA SchLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex*. *Nature Genetics*, 2013. **45**(11): p. 1392-8.
103. Gutschner, T., M. Hammerle, and S. Diederichs, *MALAT1 -- a paradigm for long noncoding RNA function in cancer*. *Journal of Molecular Medicine*, 2013. **91**(7): p. 791-801.
104. He, L. and G.J. Hannon, *MicroRNAs: small RNAs with a big role in gene regulation*. *Nature Reviews Genetics*, 2004. **5**(7): p. 522-31.
105. Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia*. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. **99**(24): p. 15524-9.
106. miRBase, *miRBase: the microRNA database*. Available online: <http://mirbase.org>. Accessed: 10/09/2015
107. Calin, G.A. and C.M. Croce, *MicroRNA signatures in human cancers*. *Nature Reviews Genetics*, 2006. **6**(11): p. 857-66.
108. Russo, J. and I.H. Russo, *Development of the human breast*. *Maturitas*, 2004. **49**(1): p. 2-15.
109. Warri, A.M., N.M. Saarinen, and S.I. Makela, *Can modulation of mammary gland development by dietary factors support breast cancer prevention?* *Hormone Research in Paediatrics*, 2007. **68**(5): p. 248-60.

110. IARC, *World Cancer Report 2014*. 1st ed. ed. 2014, Lyon, France: International Agency for Research on Cancer (IARC).
111. Ferlay, J., et al., *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2013, International Agency for Research on Cancer: Lyon, France.
112. Grayson, M., *Breast cancer*. *Nature*, 2012. **485**(7400): p. S49-S49.
113. National Cancer Institute (Bethesda MD USA), *SEER Cancer Statistics Factsheets: All Cancer Sites*. 2014.
114. Rosner, B., G.A. Colditz, and W.C. Willett, *Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study*. *American Journal of Epidemiology*, 1994. **139**(8): p. 819-35.
115. Economopoulou, P., G. Dimitriadis, and A. Psyrris, *Beyond BRCA: new hereditary breast cancer susceptibility genes*. *Cancer Treatment Reviews*, 2015. **41**(1): p. 1-8.
116. Wiechmann, L. and H.M. Kuerer, *The molecular journey from ductal carcinoma in situ to invasive breast cancer*. *Cancer*, 2008. **112**(10): p. 2130-42.
117. Allred, D.C., et al., *Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution*. *Clinical Cancer Research*, 2008. **14**(2): p. 370-8.
118. Bertos, N.R. and M. Park, *Breast cancer - one term, many entities?* *Journal of Clinical Investigation*, 2011. **121**(10): p. 3789-96.
119. Li, C.I., D.J. Uribe, and J.R. Daling, *Clinical characteristics of different histologic types of breast cancer*. *British Journal of Cancer*, 2005. **93**(9): p. 1046-52.
120. Weigelt, B. and J.S. Reis-Filho, *Histological and molecular types of breast cancer: is there a unifying taxonomy?* *Nature Reviews Clinical Oncology*, 2009. **6**(12): p. 718-30.
121. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets*. *Proceedings of the National Academy of Sciences of the United States of America*, 2003. **100**(14): p. 8418-23.
122. Perou, C.M. and A.L. Borresen-Dale, *Systems biology and genomics of breast cancer*. *Cold Spring Harbor Perspectives in Biology*, 2011. **3**(2).
123. Ziegler, R.G., et al., *Migration Patterns and Breast Cancer Risk in Asian-American Women*. *Journal of the National Cancer Institute*, 1993. **85**(22): p. 1819-1827.
124. Tham, D.M., C.D. Gardner, and W.L. Haskell, *Potential Health Benefits of Dietary Phytoestrogens: A Review of the Clinical, Epidemiological, and Mechanistic Evidence*. *The Journal of Clinical Endocrinology & Metabolism*, 1998. **83**(7): p. 2223-2235.
125. Mori, M., et al., *Traditional Japanese diet and prostate cancer*. *Molecular Nutrition & Food Research*, 2009. **53**(2): p. 191-200.
126. Hilakivi-Clarke, L., J.E. Andrade, and W. Helferich, *Is Soy Consumption Good or Bad for the Breast?* *The Journal of Nutrition*, 2010. **140**(12): p. 2326S-2334S.
127. Messina, M., W. McCaskill-Stevens, and J.W. Lampe, *Addressing the Soy and Breast Cancer Relationship: Review, Commentary, and Workshop Proceedings*. *Journal of the National Cancer Institute*, 2006. **98**(18): p. 1275-1284.
128. Lee, S.-A., et al., *Adolescent and adult soy food intake and breast cancer risk: results from the Shanghai Women's Health Study*. *The American Journal of Clinical Nutrition*, 2009. **89**(6): p. 1920-1926.

129. Iwasaki, M., et al., *Plasma Isoflavone Level and Subsequent Risk of Breast Cancer Among Japanese Women: A Nested Case-Control Study From the Japan Public Health Center-Based Prospective Study Group*. *Journal of Clinical Oncology*, 2008. **26**(10): p. 1677-1683.
130. Travis, R.C., et al., *A prospective study of vegetarianism and isoflavone intake in relation to breast cancer risk in British women*. *International Journal of Cancer*, 2008. **122**(3): p. 705-710.
131. Trock, B.J., L. Hilakivi-Clarke, and R. Clarke, *Meta-analysis of soy intake and breast cancer risk*. *Journal of the National Cancer Institute*, 2006. **98**(7): p. 459-71.
132. Ju, Y.H., et al., *Physiological concentrations of dietary genistein dose-dependently stimulate growth of estrogen-dependent human breast cancer (MCF-7) tumors implanted in athymic nude mice*. *The Journal of Nutrition*, 2001. **131**(11): p. 2957-62.
133. Allred, C.D., et al., *Soy processing influences growth of estrogen-dependent breast cancer tumors*. *Carcinogenesis*, 2004. **25**(9): p. 1649-57.
134. Liu, Y., et al., *Isoflavones in soy flour diet have different effects on whole-genome expression patterns than purified isoflavone mix in human MCF-7 breast tumors in ovariectomized athymic nude mice*. *Molecular Nutrition & Food Research*, 2015. **59**(8): p. 1419-30.
135. Tsai, C.Y., et al., *Effect of soy saponin on the growth of human colon cancer cells*. *World Journal of Gastroenterology*, 2010. **16**(27): p. 3371-6.
136. Velasquez, M.T. and S.J. Bhatena, *Role of dietary soy protein in obesity*. *International Journal of Medical Sciences*, 2007. **4**(2): p. 72-82.
137. Friedman, M. and D.L. Brandon, *Nutritional and health benefits of soy proteins*. *Journal of Agricultural and Food Chemistry*, 2001. **49**(3): p. 1069-86.
138. Su, Y., et al., *Expression profiling of rat mammary epithelial cells reveals candidate signaling pathways in dietary protection from mammary tumors*. *Physiological Genomics*, 2007. **30**(1): p. 8-16.
139. Morimoto, Y., et al., *Dietary isoflavone intake is not statistically significantly associated with breast cancer risk in the Multiethnic Cohort*. *British Journal of Nutrition*, 2014. **112**(6): p. 976-83.
140. Warri, A., et al., *The role of early life genistein exposures in modifying breast cancer risk*. *British Journal of Cancer*, 2008. **98**(9): p. 1485-93.
141. De Assis, S. and L. Hilakivi-Clarke, *Timing of Dietary Estrogenic Exposures and Breast Cancer Risk*. *Annals of the New York Academy of Sciences*, 2006. **1089**(1): p. 14-35.
142. Lamartiniere, C.A., *Timing of exposure and mammary cancer risk*. *Journal of Mammary Gland Biology and Neoplasia*, 2002. **7**(1): p. 67-76.
143. McMichael-Phillips, D.F., et al., *Effects of soy-protein supplementation on epithelial proliferation in the histologically normal human breast*. *The American Journal of Clinical Nutrition*, 1998. **68**(6 Suppl): p. 1431S-1435S.
144. Hilakivi-Clarke, L., *Nutritional modulation of terminal end buds: its relevance to breast cancer prevention*. *Current Cancer Drug Target*, 2007. **7**: p. 465-474.
145. Widschwendter, M. and P.A. Jones, *DNA methylation and breast carcinogenesis*. *Oncogene*, 2002. **21**(35): p. 5462-82.
146. Park, S.Y., et al., *Promoter CpG island hypermethylation during breast cancer progression*. *Virchows Archive*, 2011. **458**(1): p. 73-84.

147. Gascard, P., et al., *Epigenetic and transcriptional determinants of the human breast*. Nature Communications, 2015. **6**: p. 6351.
148. Lin, C.Y., et al., *Transcriptional amplification in tumor cells with elevated c-Myc*. Cell, 2012. **151**(1): p. 56-67.
149. Stefansson, O.A., et al., *A DNA methylation-based definition of biologically distinct breast cancer subtypes*. Molecular Oncology, 2015. **9**(3): p. 555-68.
150. Gyorffy, B., et al., *Aberrant DNA methylation impacts gene expression and prognosis in breast cancer subtypes*. International Journal of Cancer, 2015.
151. Szyf, M., *DNA methylation signatures for breast cancer classification and prognosis*. Genome Medicine, 2012. **4**(3): p. 26.
152. King-Batoon, A., J.M. Leszczynska, and C.B. Klein, *Modulation of gene methylation by genistein or lycopene in breast cancer cells*. Environmental and Molecular Mutagenesis, 2008. **49**(1): p. 36-45.
153. Bosviel, R., et al., *Can soy phytoestrogens decrease DNA methylation in BRCA1 and BRCA2 oncosuppressor genes in breast cancer?* OMICS, 2012. **16**(5): p. 235-44.
154. Bosviel, R., et al., *Epigenetic modulation of BRCA1 and BRCA2 gene expression by equol in breast cancer cell lines*. British Journal of Cancer, 2012. **108**(7): p. 1187-93.
155. Li, Y., et al., *Epigenetic regulation of multiple tumor-related genes leads to suppression of breast tumorigenesis by dietary genistein*. PLoS One, 2013. **8**(1): p. e54369.
156. Xie, Q., et al., *Genistein inhibits DNA methylation and increases expression of tumor suppressor genes in human breast cancer cells*. Genes, Chromosomes and Cancer, 2014. **53**(5): p. 422-431.
157. Li, Y., et al., *Genistein depletes telomerase activity through cross-talk between genetic and epigenetic mechanisms*. International Journal of Cancer, 2009. **125**(2): p. 286-96.
158. Li, Y., et al., *Epigenetic reactivation of estrogen receptor-alpha (ERalpha) by genistein enhances hormonal therapy sensitivity in ERalpha-negative breast cancer*. Molecular Cancer, 2013. **12**(1): p. 9.
159. Qin, W., et al., *Soy isoflavones have an antiestrogenic effect and alter mammary promoter hypermethylation in healthy premenopausal women*. Nutrition and Cancer, 2009. **61**(2): p. 238-44.
160. Blei, T., et al., *Dose-dependent effects of isoflavone exposure during early lifetime on the rat mammary gland: Studies on estrogen sensitivity, isoflavone metabolism, and DNA methylation*. Molecular Nutrition & Food Research, 2015. **59**(2): p. 270-83.
161. Turan, V.K., et al., *The effects of steroidal estrogens in ACI rat mammary carcinogenesis: 17beta-estradiol, 2-hydroxyestradiol, 4-hydroxyestradiol, 16alpha-hydroxyestradiol, and 4-hydroxyestrone*. The Journal of Endocrinology, 2004. **183**(1): p. 91-9.
162. Dennison, K.L., et al., *Development and characterization of a novel rat model of estrogen-induced mammary cancer*. Endocrine-Related Cancer, 2015. **22**(2): p. 239-48.
163. Blank, E.W., et al., *Both ovarian hormones estrogen and progesterone are necessary for hormonal mammary carcinogenesis in ovariectomized ACI rats*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(9): p. 3527-32.
164. Singh, B., N.K. Bhat, and H.K. Bhat, *Partial inhibition of estrogen-induced mammary carcinogenesis in rats by tamoxifen: balance between oxidant stress and estrogen responsiveness*. PLoS One, 2011. **6**(9): p. e25125.

165. Day, J.K., et al., *Genistein alters methylation patterns in mice*. The Journal of Nutrition, 2002. **132**(8 Suppl): p. 2419S-2423S.
166. Gould, K.A., et al., *Genetic determination of susceptibility to estrogen-induced mammary cancer in the ACI rat: mapping of Emca1 and Emca2 to chromosomes 5 and 18*. Genetics, 2004. **168**(4): p. 2113-25.
167. Colletti, J.A., 2nd, et al., *Validation of six genetic determinants of susceptibility to estrogen-induced mammary cancer in the rat and assessment of their relevance to breast cancer risk in humans*. G3 (Bethesda), 2014. **4**(8): p. 1385-94.
168. Kutanzi, K.R., I. Koturbash, and O. Kovalchuk, *Reversibility of pre-malignant estrogen-induced epigenetic changes*. Cell Cycle, 2010. **9**(15): p. 3078-84.
169. Weischenfeldt, J., et al., *Integrative Genomic Analyses Reveal an Androgen-Driven Somatic Alteration Landscape in Early-Onset Prostate Cancer*. Cancer Cell, 2013. **23**(2): p. 159-170.
170. Sonnet, M., et al., *Enrichment of methylated DNA by methyl-CpG immunoprecipitation*. Methods in Molecular Biology, 2013. **971**: p. 201-12.
171. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-2079.
172. Heinz, S., et al., *Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities*. Molecular Cell, 2010. **38**(4): p. 576-589.
173. Ronis, M.J.J., et al., *Mammary Gland Morphology and Gene Expression Differ in Female Rats Treated with 17 β -Estradiol or Fed Soy Protein Isolate*. Endocrinology, 2012. **153**(12): p. 6021-6032.
174. Schieber, J., *Illumina TruSeq DNA Adapters De-Mystified 2011*. Available from: http://tucf-genomics.tufts.edu/documents/protocols/TUCF_Understanding_Illumina_TruSeq_Adapters.pdf
175. Xi, Y. and W. Li, *BSMAP: whole genome bisulfite sequence MAPPING program*. BMC Bioinformatics, 2009. **10**: p. 232.
176. Bock, C., et al., *Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines*. Cell, 2011. **144**(3): p. 439-52.
177. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads*. Nature Methods, 2014. **11**(11): p. 1138-1140.
178. Livak, K.J. and T.D. Schmittgen, *Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method*. Methods, 2001. **25**(4): p. 402-408.
179. Molzberger, A.F., et al., *In utero and postnatal exposure to isoflavones results in a reduced responsivity of the mammary gland towards estradiol*. Molecular Nutrition & Food Research, 2012. **56**(3): p. 399-409.
180. Helle, J., et al., *Assessment of the proliferative capacity of the flavanones 8-prenylnaringenin, 6-(1,1-dimethylallyl)naringenin and naringenin in MCF-7 cells and the rat mammary gland*. Molecular and Cellular Endocrinology, 2014. **392**(1-2): p. 125-35.
181. Link, A., et al., *Curcumin modulates DNA methylation in colorectal cancer cells*. PLoS One, 2013. **8**(2): p. e57709.
182. Esteller, M., *Cancer epigenomics: DNA methylomes and histone-modification maps*. Nature Reviews Genetics, 2007. **8**(4): p. 286-98.

183. Chandan, V.S., et al., *Globular hepatic amyloid is highly sensitive and specific for LECT2 amyloidosis*. The American Journal of Surgical Pathology, 2015. **39**(4): p. 558-64.
184. Andres, S.A., et al., *Interaction between smoking history and gene expression levels impacts survival of breast cancer patients*. Breast Cancer Research and Treatment, 2015. **152**(3): p. 545-56.
185. Wang, H., et al., *RITA inhibits growth of human hepatocellular carcinoma through induction of apoptosis*. Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics, 2013. **20**(10): p. 437-45.
186. Wang, H., et al., *RBP-J-interacting and tubulin-associated protein induces apoptosis and cell cycle arrest in human hepatocellular carcinoma by activating the p53-Fbxw7 pathway*. Biochemical and Biophysical Research Communications, 2014. **454**(1): p. 71-7.
187. Komatsu, A., et al., *Identification of novel deletion polymorphisms in breast cancer*. International Journal of Oncology, 2008. **33**(2): p. 261-70.
188. Wang, J. and C. Abate-Shen, *The MSX1 homeoprotein recruits G9a methyltransferase to repressed target genes in myoblast cells*. PLoS One, 2012. **7**(5): p. e37647.
189. Godoy, P., et al., *Interferon-inducible guanylate binding protein (GBP2) is associated with better prognosis in breast cancer and indicates an efficient T cell response*. Breast Cancer, 2014. **21**(4): p. 491-9.
190. Ren, G.F., et al., *Prognostic impact of NDRG2 and NDRG3 in prostate cancer patients undergoing radical prostatectomy*. Histology and Histopathology, 2014. **29**(4): p. 535-42.
191. Melander, M.C., et al., *The collagen receptor uPARAP/Endo180 in tissue degradation and cancer (Review)*. International Journal of Oncology, 2015. **47**(4): p. 1177-88.
192. Khan, S., et al., *MicroRNA-10a is reduced in breast cancer and regulated in part through retinoic acid*. BMC Cancer, 2015. **15**: p. 345.
193. Traving, C. and R. Schauer, *Structure, function and metabolism of sialic acids*. Cellular and Molecular Life Sciences, 1998. **54**(12): p. 1330-49.
194. Ding, L., et al., *Association of cellular and molecular responses in the rat mammary gland to 17beta-estradiol with susceptibility to mammary cancer*. BMC Cancer, 2013. **13**: p. 573.
195. Kulis, M., et al., *Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer*. Biochimica et Biophysica Acta, 2013. **1829**(11): p. 1161-74.
196. Rauscher, G.H., et al., *Exploring DNA methylation changes in promoter, intragenic, and intergenic regions as early and late events in breast cancer formation*. BMC Cancer, 2015. **15**: p. 816.
197. Hovestadt, V., et al., *Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing*. Nature, 2014.
198. Tost, J. and I.G. Gut, *DNA methylation analysis by pyrosequencing*. Nature Protocols, 2007. **2**(9): p. 2265-75.
199. Brait, M., et al., *Association between lifestyle factors and CpG island methylation in a cancer-free population*. Cancer Epidemiology Biomarkers and Prevention, 2009. **18**(11): p. 2984-91.
200. Casati, L., et al., *Endocrine disruptors: the new players able to affect the epigenome*. Frontiers in Cell and Developmental Biology, 2015. **3**: p. 37.
201. Lee, H.S., *Impact of Maternal Diet on the Epigenome during In Utero Life and the Developmental Programming of Diseases in Childhood and Adulthood*. Nutrients, 2015. **7**(11): p. 9492-507.

202. Sonnet, M., et al., *Early aberrant DNA methylation events in a mouse model of acute myeloid leukemia*. *Genome Medicine*, 2014. **6**(4): p. 34.
203. Martinez-Arguelles, D.B., S. Lee, and V. Papadopoulos, *In silico analysis identifies novel restriction enzyme combinations that expand reduced representation bisulfite sequencing CpG coverage*. *BMC Research Notes*, 2014. **7**: p. 534.
204. Liu, Y., et al., *Base-resolution maps of 5-methylcytosine and 5-hydroxymethylcytosine in Dahl S rats: effect of salt and genomic sequence*. *Hypertension*, 2014. **63**(4): p. 827-38.
205. Akalin, A., et al., *Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia*. *PLoS Genetics*, 2012. **8**(6): p. e1002781.
206. Watson, C.T., et al., *Genome-Wide DNA Methylation Profiling Reveals Epigenetic Changes in the Rat Nucleus Accumbens Associated With Cross-Generational Effects of Adolescent THC Exposure*. *Neuropsychopharmacology*, 2015. **40**(13): p. 2993-3005.
207. Tomazou, E.M., et al., *Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1*. *Cell Reports*, 2015. **10**(7): p. 1082-95.
208. Petropoulos, S., et al., *Gestational Diabetes Alters Offspring DNA Methylation Profiles in Human and Rat: Identification of Key Pathways Involved in Endocrine System Disorders, Insulin Signaling, Diabetes Signaling, and ILK Signaling*. *Endocrinology*, 2015. **156**(6): p. 2222-38.
209. Bird, A., *DNA methylation patterns and epigenetic memory*. *Genes and Development*, 2002. **16**(1): p. 6-21.
210. Sandovici, I., et al., *Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets*. *Proceedings of the National Academy of Sciences of the United States of America*, 2011. **108**(13): p. 5449-54.
211. Heyn, H., et al., *Distinct DNA methylomes of newborns and centenarians*. *Proceedings of the National Academy of Sciences of the United States of America*, 2012. **109**(26): p. 10522-7.
212. Kovalchuk, O., et al., *Estrogen-induced rat breast carcinogenesis is characterized by alterations in DNA methylation, histone modifications and aberrant microRNA expression*. *Cell Cycle*, 2007. **6**(16): p. 2010-8.
213. Esteller, M. and G. Almouzni, *How epigenetics integrates nuclear functions. Workshop on epigenetics and chromatin: transcriptional regulation and beyond*. *EMBO Reports*, 2005. **6**(7): p. 624-8.
214. Wang, Y., et al., *A novel active endogenous retrovirus family contributes to genome variability in rat inbred strains*. *Genome Research*, 2010. **20**(1): p. 19-27.
215. Sati, S., et al., *High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region*. *PLoS One*, 2012. **7**(2): p. e31621.
216. Anastasiadou, C., et al., *Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers*. *DNA and Cell Biology*, 2011. **30**(5): p. 267-75.
217. Wan, J., et al., *Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs*. *Nucleic Acids Research*, 2013. **41**(18): p. 8503-14.
218. Plass, C. and P.D. Soloway, *DNA methylation, imprinting and cancer*. *European Journal of Human Genetics*, 2002. **10**(1): p. 6-16.

219. Wu, Q., et al., *Exposure of mouse preimplantation embryos to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) alters the methylation status of imprinted genes H19 and Igf2*. *Biology of Reproduction*, 2004. **70**(6): p. 1790-7.
220. Tanaka, K. and A. Okamoto, *Degradation of DNA by bisulfite treatment*. *Bioorganic and Medicinal Chemistry Letters*, 2007. **17**(7): p. 1912-5.
221. Furst, R.W., et al., *A differentially methylated single CpG-site is correlated with estrogen receptor alpha transcription*. *The Journal of Steroid Biochemistry and Molecular Biology*, 2012. **130**(1-2): p. 96-104.
222. Huynh, J.L., et al., *Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains*. *Nature Neuroscience*, 2014. **17**(1): p. 121-30.
223. Kivioja, T., et al., *Counting absolute numbers of molecules using unique molecular identifiers*. *Nature Methods*, 2012. **9**(1): p. 72-4.
224. Shipony, Z., et al., *Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells*. *Nature*, 2014. **513**(7516): p. 115-9.
225. Kaur, J.S., *Migration patterns and breast carcinoma*. *Cancer*, 2000. **88**(5 Suppl): p. 1203-6.
226. Kushner, P.J., et al., *Estrogen receptor pathways to AP-1*. *The Journal of Steroid Biochemistry and Molecular Biology*, 2000. **74**(5): p. 311-7.
227. Shen, Q., et al., *The AP-1 transcription factor regulates breast cancer cell growth via cyclins and E2F factors*. *Oncogene*, 2008. **27**(3): p. 366-77.
228. Gutierrez-Hartmann, A., D.L. Duval, and A.P. Bradford, *ETS transcription factors in endocrine systems*. *Trends in Endocrinology and Metabolism*, 2007. **18**(4): p. 150-8.
229. Sharrocks, A.D., *The ETS-domain transcription factor family*. *Nature Reviews Molecular Cell Biology*, 2001. **2**(11): p. 827-37.
230. Messina, M., C. Nagata, and A.H. Wu, *Estimated Asian adult soy protein and isoflavone intakes*. *Nutrition and Cancer*, 2006. **55**(1): p. 1-12.
231. Schneider, K., J. Oltmanns, and M. Hassauer, *Allometric principles for interspecies extrapolation in toxicological risk assessment--empirical investigations*. *Regulatory Toxicology and Pharmacology*, 2004. **39**(3): p. 334-47.
232. Rucker, R. and D. Storms, *Interspecies comparisons of micronutrient requirements: metabolic vs. absolute body size*. *The Journal of Nutrition*, 2002. **132**(10): p. 2999-3000.
233. Setchell, K.D. and E. Lydeking-Olsen, *Dietary phytoestrogens and their effect on bone: evidence from in vitro and in vivo, human observational, and dietary intervention studies*. *Am J Clin Nutr*, 2003. **78**(3 Suppl): p. 593S-609S.
234. Dinsdale, E.C. and W.E. Ward, *Early exposure to soy isoflavones and effects on reproductive health: a review of human and animal studies*. *Nutrients*, 2010. **2**(11): p. 1156-87.
235. Molzberger, A.F., et al., *Proliferative and estrogenic sensitivity of the mammary gland are modulated by isoflavones during distinct periods of adolescence*. *Archives of Toxicology*, 2013. **87**(6): p. 1129-40.
236. National Toxicology, P., *Multigenerational reproductive study of genistein (Cas No. 446-72-0) in Sprague-Dawley rats (feed study)*. National Toxicology Program technical report series, 2008(539): p. 1-266.

237. Guerrero-Bosagna, C.M., et al., *Epigenetic and phenotypic changes result from a continuous pre and post natal dietary exposure to phytoestrogens in an experimental population of mice*. BMC Physiology, 2008. **8**: p. 17.
238. Adgent, M.A., et al., *Early-life soy exposure and age at menarche*. Paediatric and Perinatal Epidemiology, 2012. **26**(2): p. 163-75.
239. Kim, S.H. and M.J. Park, *Effects of phytoestrogen on sexual development*. Korean Journal of Pediatrics, 2012. **55**(8): p. 265-71.
240. Golub, M.S., et al., *Public health implications of altered puberty timing*. Pediatrics, 2008. **121** Suppl 3: p. S218-30.
241. Cheng, G., et al., *Relation of isoflavones and fiber intake in childhood to the timing of puberty*. The American Journal of Clinical Nutrition, 2010. **92**(3): p. 556-64.
242. D'Aloisio, A.A., et al., *Prenatal and infant exposures and age at menarche*. Epidemiology, 2013. **24**(2): p. 277-84.
243. Cheng, G., et al., *Beyond overweight: nutrition as an important lifestyle factor influencing timing of puberty*. Nutrition Reviews, 2012. **70**(3): p. 133-52.
244. Faber, K.A. and C.L. Hughes, Jr., *Dose-response characteristics of neonatal exposure to genistein on pituitary responsiveness to gonadotropin releasing hormone and volume of the sexually dimorphic nucleus of the preoptic area (SDN-POA) in postpubertal castrated female rats*. Reproductive Toxicology, 1993. **7**(1): p. 35-9.
245. Warri, A., et al., *The role of early life genistein exposures in modifying breast cancer risk*. British Journal of Cancer, 2008. **98**(9): p. 1485-93.
246. Zierau, O., et al., *Time dependency of uterine effects of naringenin type phytoestrogens in vivo*. Molecular and Cellular Endocrinology, 2008. **294**(1-2): p. 92-9.
247. Qian, P., et al., *Loss of SNAIL regulated miR-128-2 on chromosome 3p22.3 targets multiple stem cell factors to promote transformation of mammary epithelial cells*. Cancer Research, 2012. **72**(22): p. 6036-50.
248. Shi, Y., et al., *PTEN at a glance*. Journal of Cell Science, 2012. **125**(Pt 20): p. 4687-92.
249. Ma, X., et al., *Claudin-4 controls the proliferation, apoptosis, migration and in vivo growth of MCF-7 breast cancer cells*. Oncology Reports, 2015. **34**(2): p. 681-90.
250. Ma, F., et al., *A CLDN1-negative phenotype predicts poor prognosis in triple-negative breast cancer*. PLoS One, 2014. **9**(11): p. e112765.
251. Lapidus, R.G., S.J. Nass, and N.E. Davidson, *The loss of estrogen and progesterone receptor gene expression in human breast cancer*. Journal of mammary gland biology and neoplasia, 1998. **3**(1): p. 85-94.
252. Hatsumi, T. and Y. Yamamuro, *Downregulation of estrogen receptor gene expression by exogenous 17beta-estradiol in the mammary glands of lactating mice*. Experimental Biology and Medicine 2006. **231**(3): p. 311-6.
253. Reynolds, A.R., *Potential relevance of bell-shaped and u-shaped dose-responses for the therapeutic targeting of angiogenesis in cancer*. Dose Response, 2009. **8**(3): p. 253-84.

-
254. Parris, G.E., *A Hypothesis Concerning the Biphasic Dose-response of Tumors to Angiostatin and Endostatin*. Dose Response, 2015. **13**(2).
255. Calabrese, E.J., J.W. Staudenmayer, and E.J. Stanek, *Drug development and hormesis: changing conceptual understanding of the dose response creates new challenges and opportunities for more effective drugs*. Current Opinion in Drug Discovery and Development, 2006. **9**(1): p. 117-23.
256. Messina, M.J. and C.L. Loprinzi, *Soy for breast cancer survivors: a critical review of the literature*. The Journal of Nutrition, 2001. **131**(11 Suppl): p. 3095S-108S.
257. Iwasaki, M., et al., *Plasma isoflavone level and subsequent risk of breast cancer among Japanese women: a nested case-control study from the Japan Public Health Center-based prospective study group*. Journal of Clinical Oncology, 2008. **26**(10): p. 1677-83.
258. Foster, W.G., et al., *Detection of phytoestrogens in samples of second trimester human amniotic fluid*. Toxicology Letters, 2002. **129**(3): p. 199-205.
259. Doerge, D.R., et al., *Lactational transfer of the soy isoflavone, genistein, in Sprague-Dawley rats consuming dietary genistein*. Reproductive Toxicology, 2006. **21**(3): p. 307-12.
260. Wietzke, J.A. and J. Welsh, *Phytoestrogen regulation of a Vitamin D3 receptor promoter and 1,25-dihydroxyvitamin D3 actions in human breast cancer cells*. The Journal of Steroid Biochemistry and Molecular Biology, 2003. **84**(2-3): p. 149-57.
261. Thomsen, A.R., et al., *Estrogenic effect of soy isoflavones on mammary gland morphogenesis and gene expression profile*. Toxicology Sciences, 2006. **93**(2): p. 357-68.
262. Wanders, D., E.C. Graff, and R.L. Judd, *Effects of high fat diet on GPR109A and GPR81 gene expression*. Biochemical and Biophysical Research Communications, 2012. **425**(2): p. 278-83.
263. Lamartiniere, C.A., *Timing of exposure and mammary cancer risk*. J Mammary Gland Biol Neoplasia, 2002. **7**(1): p. 67-76.
264. Wang, J., et al., *Proteomic discovery of genistein action in the rat mammary gland*. Journal of Proteome Research, 2011. **10**(4): p. 1621-31.
265. Aparecida Santos, M., et al., *Effects of early and late treatment with soy isoflavones in the mammary gland of ovariectomized rats*. Climacteric, 2015: p. 1-8.
266. Abd-Elazeem, M.A. and M.A. Abd-Elazeem, *Claudin 4 expression in triple-negative breast cancer: correlation with androgen receptors and Ki-67 expression*. Annals of Diagnostic Pathology, 2015. **19**(1): p. 37-42.
267. Shike, M., et al., *The effects of soy supplementation on gene expression in breast cancer: a randomized placebo-controlled study*. Journal of the National Cancer Institute, 2014. **106**(9).
268. Jordan, V.C., *Avoiding the bad and enhancing the good of soy supplements in breast cancer*. Journal of the National Cancer Institute, 2014. **106**(9).
269. Wuttke, W., H. Jarry, and D. Seidlova-Wuttke, *Isoflavones--safe food additives or dangerous drugs?* Ageing Research Reviews, 2007. **6**(2): p. 150-88.
270. Fritz, W.A., et al., *Dietary genistein: perinatal mammary cancer prevention, bioavailability and toxicity testing in the rat*. Carcinogenesis, 1998. **19**(12): p. 2151-8.
271. Crambert, G. and K. Geering, *FXYP proteins: new tissue-specific regulators of the ubiquitous Na,K-ATPase*. Science's Signal Transduction Knowledge Environment, 2003. **2003**(166): p. RE1.

-
272. La Vecchia, C., et al., *Overweight, obesity, diabetes, and risk of breast cancer: interlocking pieces of the puzzle*. *Oncologist*, 2011. **16**(6): p. 726-9.
273. Mohan, S., et al., *Structure of Shroom domain 2 reveals a three-segmented coiled-coil required for dimerization, Rock binding, and apical constriction*. *Molecular Biology of the Cell*, 2012. **23**(11): p. 2131-42.
274. Nagarajan, S., et al., *H4K12ac is regulated by estrogen receptor-alpha and is associated with BRD4 function and inducible transcription*. *Oncotarget*, 2015. **6**(9): p. 7305-17.
275. Chen, K., et al., *Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer*. *Cell Research*, 2016. **26**(1): p. 103-18.
276. Branco, M.R., G. Ficz, and W. Reik, *Uncovering the role of 5-hydroxymethylcytosine in the epigenome*. *Nature Reviews Genetics*, 2012. **13**(1): p. 7-13.
277. Hennighausen, L. and G.W. Robinson, *Information networks in the mammary gland*. *Nature Reviews Molecular Cell Biology*, 2005. **6**(9): p. 715-25.
278. Vargo-Gogola, T. and J.M. Rosen, *Modelling breast cancer: one size does not fit all*. *Nature Reviews Cancer*, 2007. **7**(9): p. 659-72.
279. Turan, V.K., et al., *The effects of steroidal estrogens in ACI rat mammary carcinogenesis: 17beta-estradiol, 2-hydroxyestradiol, 4-hydroxyestradiol, 16alpha-hydroxyestradiol, and 4-hydroxyestrone*. *J Endocrinology*, 2004. **183**(1): p. 91-9.
280. Barker, D.J. and P.M. Clark, *Fetal undernutrition and disease in later life*. *Reviews of Reproduction*, 1997. **2**(2): p. 105-12.
281. Wadhwa, P.D., et al., *Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms*. *Seminars in Reproductive Medicine*, 2009. **27**(5): p. 358-68.
282. Walker, C.L. and S.M. Ho, *Developmental reprogramming of cancer susceptibility*. *Nature Reviews Cancer*, 2012. **12**(7): p. 479-86.
283. Betancourt, A.M., et al., *Altered carcinogenesis and proteome in mammary glands of rats after prepubertal exposures to the hormonally active chemicals bisphenol a and genistein*. *The Journal of Nutrition*, 2012. **142**(7): p. 1382S-8S.
284. Watson, K.L., et al., *High levels of dietary soy decrease mammary tumor latency and increase incidence in MTB-IGFIR transgenic mice*. *BMC Cancer*, 2015. **15**: p. 37.
285. Ju, Y.H., et al., *Physiological concentrations of dietary genistein dose-dependently stimulate growth of estrogen-dependent human breast cancer (MCF-7) tumors implanted in athymic nude mice*. *J Nutr*, 2001. **131**(11): p. 2957-62.
286. Santell, R.C., N. Kieu, and W.G. Helferich, *Genistein inhibits growth of estrogen-independent human breast cancer cells in culture but not in athymic mice*. *The Journal of Nutrition*, 2000. **130**(7): p. 1665-9.
287. Kutanzi, K.R., et al., *Imbalance between apoptosis and cell proliferation during early stages of mammary gland carcinogenesis in ACI rats*. *Mutation Research*, 2010. **694**(1-2): p. 1-6.
288. Gascard, P., et al., *Epigenetic and transcriptional determinants of the human breast*. *Nature Communication*, 2015. **6**: p. 6351.
289. Schwartzman, O. and A. Tanay, *Single-cell epigenomics: techniques and emerging applications*. *Nature Reviews Genetics*, 2015. **16**(12): p. 716-26.

-
290. Neri, F., et al., *TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway*. *Oncogene*, 2015. **34**(32): p. 4168-76.
 291. Abouzeid, H., et al., *RAX and anophthalmia in humans: evidence of brain anomalies*. *Molecular Vision*, 2012. **18**: p. 1449-56.
 292. Ha Thi, H.T., et al., *Smad7 Modulates Epidermal Growth Factor Receptor Turnover through Sequestration of c-Cbl*. *Molecular and Cellular Biology*, 2015. **35**(16): p. 2841-50.
 293. Park, S.H., et al., *Itch E3 ubiquitin ligase positively regulates TGF-beta signaling to EMT via Smad7 ubiquitination*. *Molecules and Cells*, 2015. **38**(1): p. 20-5.
 294. Tanigawa, S., et al., *Jun dimerization protein 2 is a critical component of the Nrf2/MafK complex regulating the response to ROS homeostasis*. *Cell Death and Disease*, 2013. **4**: p. e921.
 295. Hwang, Y.J., et al., *MafK positively regulates NF-kappaB activity by enhancing CBP-mediated p65 acetylation*. *Scientific Reports*, 2013. **3**: p. 3242.
 296. Su, Y., et al., *Expression profiling of rat mammary epithelial cells reveals candidate signaling pathways in dietary protection from mammary tumors*. *Physiological Genomics*, 2007. **30**(1): p. 8-16.
 297. Zhang, Y., Q. Li, and H. Chen, *DNA methylation and histone modifications of Wnt genes by genistein during colon cancer development*. *Carcinogenesis*, 2013. **34**(8): p. 1756-63.
 298. Shaoqiang, C., et al., *Expression of HOXD3 correlates with shorter survival in patients with invasive breast cancer*. *Clinical and Experimental Metastasis*, 2013. **30**(2): p. 155-63.
 299. Chen, L.N., et al., *Correlation of HOXD3 promoter hypermethylation with clinical and pathologic features in screening prostate biopsies*. *Prostate*, 2014. **74**(7): p. 714-21.
 300. Bredfeldt, T.G., et al., *Xenoestrogen-induced regulation of EZH2 and histone methylation via estrogen receptor signaling to PI3K/AKT*. *Molecular Endocrinology*, 2010. **24**(5): p. 993-1006.
 301. Greathouse, K.L., et al., *Environmental estrogens differentially engage the histone methyltransferase EZH2 to increase risk of uterine tumorigenesis*. *Molecular Cancer Research*, 2012. **10**(4): p. 546-57.

8. Supplements

Supplemental table 1: MCIP-Seq statistics for HS and SF elution protocol.

Animals sequenced [number of lanes]	Elution protocol	Total reads [mio]	PF reads [mio]	PF reads [%]	Read length	PF bases [%]	PF bases [mio]	GC content [%]
12 [2]	HS	214,2	165,9	77.45	51	92.69	8,461	53.15
	HS	214,1	167,1	78.08	51	91,09	8,525	52.91
18 [4]	SF	197,8	147,3	74.45	51	93,75	7,513	46.17
	SF	206,2	135,3	65.64	51	92,57	6,902	46.12
	SF	189,9	157,1	82.70	51	93,56	8,013	49.10
	SF	187,3	159,7	85.26	51	94,18	8,155	49.13

Numbers given per lane sequenced, HS: High salt elution, SF: Single fraction elution, PF: Purity filtered bases/reads with more than 30x coverage

Supplemental table 2: Selected MClp-Seq candidates with 4-fold enrichment difference and EpiTYPER amplicon information.

	Gene Name	Refseq	Gene Description	Chr.	Strand	Start	End	Distance to TSS	Annotation	Methylation CC diet [%]
Candidates from MClp-Seq	PTEN	NM_031606	phosphatase and tensin homolog	chr1	+	258712627	258713086	61027	intron (7 of 8)	95
	Extl1	NM_001107985	exostosin-like glycosyltransferase 1	chr5	-	156387536	156387929	-1851	promoter-TSS	61
	Sirt4	NM_001107147	sirtuin 4	chr12	-	48667613	48668016	2223	intron (2 of 4)	56
	Gadd45b	NM_001008321	growth arrest and DNA-damage-inducible, beta	chr7	-	11809261	11809588	6141	intergenic	24
	Wap	NM_053751	whey acidic protein	chr14	-	80533289	80533673	3	promoter-TSS	72
	Wif1	NM_053738	Wnt inhibitory factor 1	chr7	+	63116145	63116580	38674	intron (2 of 9)	81
	Cdh13	NM_138889	cadherin 13	chr19	+	61925599	61925934	304332	intron (2 of 13)	79
	Gsn	NM_001004080	gelsolin	chr3	+	19799763	19800214	59	promoter-TSS	0.05
	Esr1	NM_012689	estrogen receptor 1	chr1	-	42666924	42667302	-72	promoter-TSS	3
	Cldn4	NM_001012022	claudin 4	chr12	-	26762009	26762384	716	promoter-TSS	23
	Aldh1L1	NM_022547	aldehyde dehydrogenase 1 family, member L1	chr4	+	186989977	186990328	-664	promoter-TSS	77
	Cyp2c6v1	NM_001013904	cytochrome P450, family 2, subfamily C, polypeptide 6, variant 1	chr1	-	154015365	154015688	9598	exon (3 of 9)	80

Chr.: Chromosome; TSS: transcription start site ; C: control diet; n.d. not determined

Supplemental table 3: Sequencing result of the puberty study of ACI rats.

Sample ID	Species	Sex	Age [PND]	Tissue	Diet	No. of lanes	Reads total [mio]	Reads aligned [mio]	Align-ment rate [%]	Total sequenced motifs [mio]	Unique sequenced motifs [mio]	Mean coverage [x]	Bisulfite conversion rate [%]	Global Methylation mean [%]
ACI_AR1-1	rat	female	21	MG	C	1	43.84	32.00	76.00	57.37	1.230	46.63	98.16	62.36
ACI_AR1-2	rat	female	21	MG	C	1	22.07	15.49	74.00	27.26	0.949	28.73	98.19	63.98
ACI_AR1-3	rat	female	21	MG	C	1	17.79	12.86	75.00	23.47	0.830	28.26	98.22	63.20
ACI_AR1-4	rat	female	21	MG	C	1	23.37	16.73	74.00	30.74	0.926	33.18	98.31	63.53
ACI_AR1-5	rat	female	21	MG	C	1	25.08	17.93	74.00	33.16	0.946	35.06	98.32	63.46
ACI_AR1-6	rat	female	21	MG	C	1	23.66	16.80	74.00	30.97	0.921	33.61	98.32	64.01
ACI_AR2-1	rat	female	21	MG	H	1	17.06	12.32	75.00	22.75	0.779	29.22	98.34	62.57
ACI_AR2-2	rat	female	21	MG	H	1	35.37	25.07	74.00	45.56	1.051	43.36	98.38	66.47
ACI_AR2-3	rat	female	21	MG	H	1	24.95	17.75	74.00	32.63	1.416	23.04	98.10	64.83
ACI_AR2-4	rat	female	21	MG	H	1	38.77	26.95	74.00	48.12	1.259	38.24	98.18	64.15
ACI_AR2-5	rat	female	21	MG	H	1	28.51	20.10	74.00	37.53	1.078	34.82	98.29	60.67
ACI_AR2-6	rat	female	21	MG	H	1	24.24	17.53	75.00	31.83	1.380	23.07	98.07	65.74
ACI_AR7-1	rat	female	50	MG	C	1	29.62	21.70	76.00	39.74	1.044	38.07	98.27	62.50
ACI_AR7-2	rat	female	50	MG	C	1	47.60	32.84	73.00	59.33	1.234	48.09	98.28	62.73
ACI_AR7-3	rat	female	50	MG	C	1	23.83	17.15	75.00	31.85	1.461	21.80	98.05	66.09
ACI_AR7-4	rat	female	50	MG	C	1	33.41	23.45	74.00	44.09	1.648	26.75	98.06	65.61
ACI_AR7-5	rat	female	50	MG	C	1	31.12	21.54	73.00	39.51	1.588	24.88	98.05	65.10
ACI_AR7-6	rat	female	50	MG	C	1	22.26	16.12	75.00	29.59	1.442	20.52	98.01	65.50
ACI_AR8-1	rat	female	50	MG	H	1	32.05	23.08	76.00	41.50	1.148	36.16	98.19	63.68
ACI_AR8-2	rat	female	50	MG	H	1	18.41	13.01	74.00	24.22	0.877	27.62	98.28	62.30
ACI_AR8-3	rat	female	50	MG	H	1	28.52	20.12	74.00	37.55	1.059	35.45	98.27	60.06
ACI_AR8-4	rat	female	50	MG	H	1	38.21	27.01	74.00	50.30	1.196	42.07	98.35	64.79
ACI_AR8-5	rat	female	50	MG	H	1	25.20	18.05	75.00	32.83	1.087	30.22	98.18	62.84
ACI_AR8-6	rat	female	50	MG	H	1	24.37	17.08	74.00	31.00	1.052	29.48	98.16	63.35

PND: post natal day, MG: mammary gland, C: control diet, H: Isoflavone enriched diet

Supplemental table 4: RRBS candidates from the puberty and the carcinogenesis study of ACI rats and EpiTYPER amplicon information.

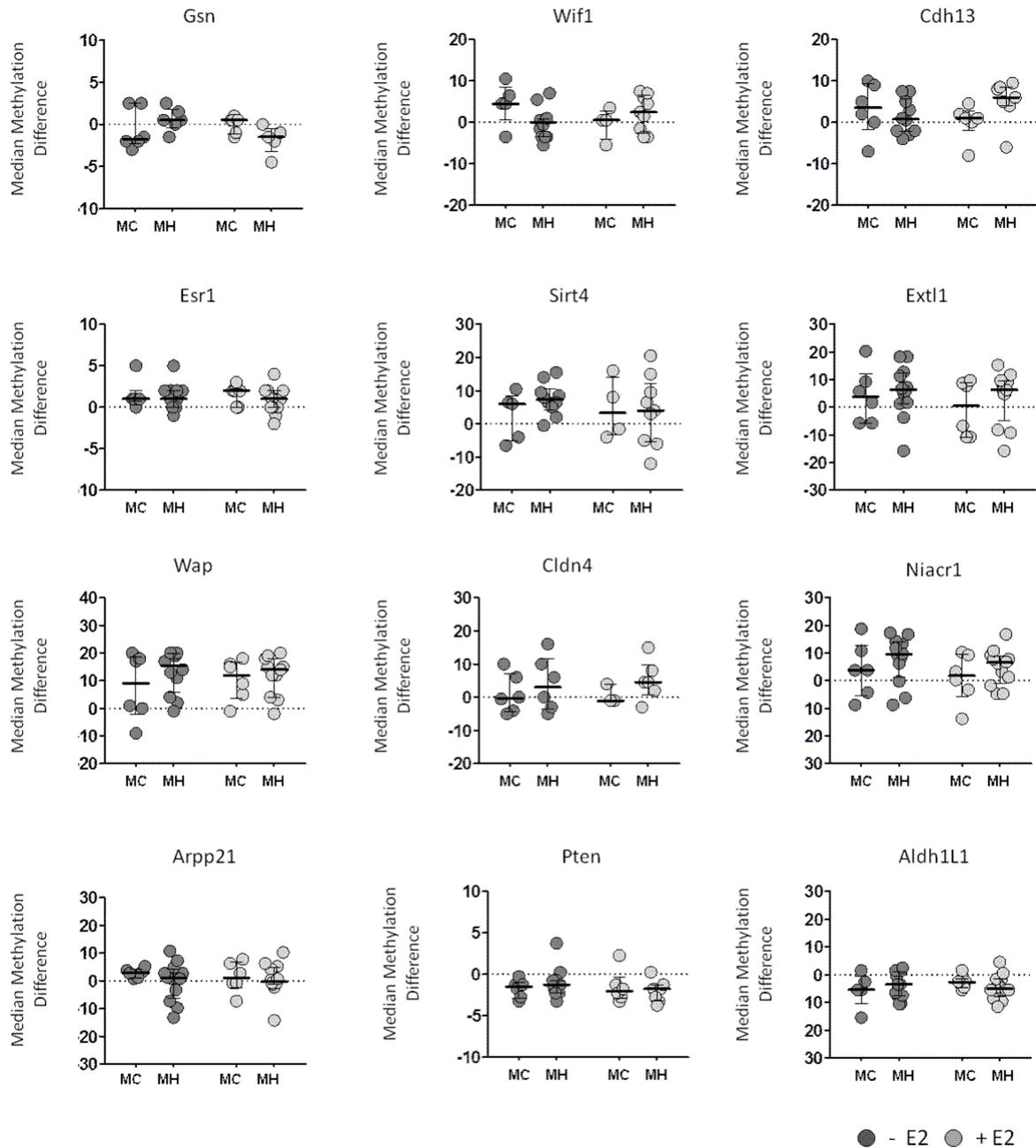
	Gene Name	Refseq	Gene description	Chr.	Strand	Start	End	Distance to TSS	Methylation [%]		PND 21 C diet	PND 285 C diet
									Annotation	CC diet		
Carcinogenesis study (PND 285)	Ndr3	NM_001013923	NDRG family member 3	3	+	158615598	158615949	1578	promoter-TSS	n.d.	n.d.	57
	Lect2	NM_001108405	leukocyte cell-derived chemotaxin 2	17	-	10664435	10664658	983	promoter-TSS	n.d.	n.d.	20
	Mrc2	NM_001024687	mannose receptor, C type 2	10	+	93277747	93278013	-1038	promoter-TSS	n.d.	n.d.	15
	Npas1	NM_001107479	neuronal PAS domain protein 1	1	+	79684310	79684665	3492	Intergenic	n.d.	n.d.	34
	Msx1	NM_031059	msh homeobox 1	14	+	77689344	77689577	-1530	promoter-TSS	n.d.	n.d.	14
	Rita	NM_001044226	RBPJ interacting and tubulin associated 1	12	-	43344158	43344363	-851	promoter-TSS	n.d.	n.d.	26
	Gbp2	NM_133624	guanylate binding protein 2, interferon-inducible	2	+	266802593	266802868	-2504	5' UTR	n.d.	n.d.	79
	Thy1	NM_012673	Thy-1 cell surface antigen	8	+	46885215	46885651	-113190	Intergenic	n.d.	n.d.	55
	Rnf8	NM_001025727	ring finger protein 8, E3 ubiquitin protein ligase	20	+	10486954	10487240	2202	intron (1 of 7)	n.d.	n.d.	55
	St3gal4	NM_203337	ST3 beta-galactoside alpha-2,3-sialyltransferase 4	8	+	36194911	36195342	138471	intron (4 of 15)	n.d.	n.d.	66
Mir10a	NR_031814	microRNA 10a	10	+	83965989	83966312	-2532	Intergenic	n.d.	n.d.	29	
Puberty study (PND 21-50)	Nr1h3	NM_031627	nuclear receptor subfamily 1, group H, member 3	3	-	86722683	86722979	1335	promoter-TSS	n.d.	55	n.d.
	Rbm42	NM_001014159	RNA binding motif protein 42	1	+	90265254	90265482	-524	promoter-TSS	n.d.	17	n.d.
	Tlr3	NM_198791	toll-like receptor 3	16	-	49739421	49739691	258	promoter-TSS	n.d.	67	n.d.
	Bpgm	NM_199382	2,3-bisphosphoglycerate mutase	4	-	61649927	61650191	822	promoter-TSS	n.d.	8	n.d.
	Niacr1	NM_181476	niacin receptor 1	12	-	40029765	40029999	-98	promoter-TSS	50	72	42
	Arpp21	NM_001135045	cAMP-regulated phosphoprotein 21	8	-	119786799	119787190	-1247	promoter-TSS	83	49	85
	Shroom2	NM_001047893	shroom family member 2	X	-	24066659	24067122	-714	promoter-TSS	n.d.	28	37
	Zfp507	NM_001106248	zinc finger protein 507	1	+	93252329	93252728	-1902	promoter-TSS	n.d.	66	62
	Fxyd2	NM_017349	FXD domain-containing ion transport regulator 2	8	-	48337077	48337170	-957	promoter-TSS	n.d.	73	46
	Mmaa	NM_001106174	methylmalonic aciduria type A protein mitochondrial	19	-	43204987	43205226	-14004	promoter-TSS	n.d.	72	85

Chr.: Chromosome; TSS: transcription start site; CC: and C diet: control diet; n.d. not determined, RRBS: Reduced representation bisulfite sequencing

Supplemental table 5: Sequencing result of the carcinogenesis study of ACI rats.

Sample ID	Species	Sex	Age [PND]	Tissue	Diet	No. of lanes	Reads total [mio]	Reads aligned [mio]	Align-ment rate [%]	Total sequenced motifs [mio]	Unique sequenced motifs [mio]	Mean coverage [x]	Bisulfite conversion rate [%]	Global Methyl. mean [%]
ACI_BA1-1	rat	Female	285	MG	C	2	60.10	45.37	75.56	95.13	1.494	63.65	98.80	56.59
ACI_BA1-3	rat	Female	285	MG	C	2	63.87	55.04	86.46	126.96	1.709	74.30	99.01	49.32
ACI_BA1-4	rat	Female	285	MG	C	2	65.55	57.04	87.16	126.55	1.680	75.34	99.01	52.41
ACI_BA1-7	rat	Female	285	MG	C	2	59.76	52.46	88.08	115.84	1.642	70.54	99.02	52.93
ACI_BA1-9	rat	Female	285	MG	C	2	81.54	68.01	83.67	162.00	1.800	89.99	98.99	48.22
ACI_BA4-2	rat	Female	285	MG	H	2	65.15	49.10	75.43	103.72	1.565	66.29	98.82	56.36
ACI_BA4-3	rat	Female	285	MG	H	2	65.75	50.88	77.45	106.57	1.570	67.88	98.82	57.36
ACI_BA4-4	rat	Female	285	MG	H	2	73.63	60.49	82.19	136.63	1.570	87.02	99.17	52.27
ACI_BA4-5	rat	Female	285	MG	H	2	57.43	46.67	81.34	97.16	1.560	62.27	98.82	57.75
ACI_BA4-6	rat	Female	285	MG	H	2	73.27	58.94	80.48	134.27	1.599	83.99	99.17	50.08
ACI_BA4-9	rat	Female	285	MG	H	2	69.52	54.16	77.98	116.74	1.586	73.63	98.82	54.22
ACI_BA3-1	rat	Female	285	T	C	2	68.94	53.75	78.09	117.16	1.654	70.86	99.02	50.37
ACI_BA3-3	rat	Female	285	T	C	2	67.85	55.69	82.21	122.77	1.650	74.42	99.04	49.88
ACI_BA3-5	rat	Female	285	T	C	2	63.87	54.48	85.33	118.10	1.498	78.82	99.19	53.6
ACI_BA3-6	rat	Female	285	T	C	2	71.85	55.59	77.50	122.01	1.664	73.32	99.03	54.89
ACI_BA3-8	rat	Female	285	T	C	2	72.64	55.20	76.02	121.53	1.541	78.88	99.19	50.54
ACI_BA3-10	rat	Female	285	T	C	2	58.96	46.22	78.65	102.64	1.607	63.87	99.01	51.49
ACI_BA6-2	rat	Female	285	T	H	2	72.67	49.68	68.40	110.55	1.498	73.79	99.18	53.26
ACI_BA6-2	rat	Female	285	T	H	2	60.27	48.07	79.89	103.47	1.538	67.27	99.01	54.14
ACI_BA6-3	rat	Female	285	T	H	2	72.45	26.60	36.75	54.51	1.256	43.41	98.83	56.38
ACI_BA6-8	rat	Female	285	T	H	2	55.93	45.54	81.47	97.13	1.379	70.45	99.18	54.62
ACI_BA6-8	rat	Female	285	T	H	2	79.74	49.12	61.70	105.30	1.568	67.15	99.02	55.64
ACI_BA6-9	rat	Female	285	T	H	2	62.04	48.90	79.08	104.06	1.609	64.68	99.01	51.99
ACI_BA6-10	rat	Female	285	T	H	2	62.12	53.19	85.90	114.28	1.599	71.47	99.02	55.07

PND: post natal day, MG: mammary gland, T: tumor, C: control diet, H: Isoflavone enriched diet



Supplemental figure 1: Summary of DNA methylation in MC and MH groups.

Each dot in the scatter plots represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Median values per group are indicated by a horizontal line, vertical lines indicate interquartile range. Median methylation of CC group is indicated as horizontal dashed line and summarized in Supplemental table 1. No statistically significant differences between MC and MH have been observed (Supplemental figure 2). Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), medium (M), and high (H) IF diets).

0.0

3.0

A

	Aldh1L1	Arpp21	Cdh13	Cldn4	Cyp2C6v1	Esr1	Extl1	Gadd45	Gsn	Niacr1	Pten	Sirt4	Wap	Wif1
CC-LL	8,5	-2,75	-1	5	8	-0,5	-9,25	5,5	2,5	-0,25	4,75	-4	-5,5	2,5
CC-MC	5,5	-3	-3,5	0,25	-3	-1	-3,75	-1	1,75	-3,75	1,5	-6	-9	-4,5
CC-MH	3,5	-1	-0,75	-3	1	-1	-6,25	-7	-0,5	-9,5	1,25	-7,2	-15,5	0
CC-CHC	-0,5	1,25	0,25	5	1,5	0	21,2	0,75	1,5	8,75	1,25	9,2	0,5	-1,5
CH-CHC	2	1	-0,75	-4	1,5	-2	4,5	-1,75	-1	-3	-5,5	-0,7	-5,5	-1
CHC-LL	9	-4	-1,25	0	6,5	-0,5	-30,5	4,75	1	-9	3,5	-13,2	-6	4
CHC-MC	6	-4,25	-3,75	-4,75	-4,5	-1	-25	-1,75	0,25	-12,5	0,25	-15,2	-9,5	-3
CHC-MH	4	-2,25	-1	-8	-0,5	-1	-27,5	-7,7	-2	-18,2	0	-16,5	-16	1,5
CH-CC	-2,5	0,25	1	9	0	2	16,7	2,5	2,5	11,7	6,7	10	6	-0,5
CH-LL	11	-3	-2	-4	8	-2,5	-26	3	0	-12	-2	-14	-11,5	3
CH-MC	8	-3,25	-4,5	-8,7	-3	-3	-20,5	-3,5	-0,75	-15,5	-5,2	-16	-15	-4
CH-MH	6	-1,25	-1,75	-12	1	-3	-23	-9,5	-3	-21,2	-5,5	-17,2	-21,5	0,5
LL-MC	-3	-0,25	-2,5	-4,75	-11	-0,5	5,5	-6,5	-0,75	-3,5	-3,25	-2	-3,5	-7
LL-MH	-5	1,75	0,25	-8	-7	-0,5	3	-12,5	-3	-9,2	-3,5	-3,2	-10	-2,5
MH-MC	2	-2	-2,75	3,25	-4	0	2,5	6	2,25	5,7	0,25	1,2	6,5	-4,5
HH-CC	-4	1	0	8	-0,5	0	8,5	0	1,5	7,5	-0,5	-1	14,5	-5,5
HH-CH	-6,5	1,25	1	17	-0,5	2	25,2	2,5	4	19,2	6,5	9	20,5	-6
HH-CHC	-4,5	2,25	0,25	13	1	0	29,7	0,75	3	16,2	0,75	8,25	15	-7
HH-LL	4,5	-1,75	-1	13	7,5	-0,5	-0,75	5,5	4	7,2	4,25	-5	9	-3
HH-MC	1,5	-2	-3,5	8,25	-3,5	-1	4,75	-1	3,25	3,7	1	-7	5,5	-10
HH-MH	-0,5	0	-0,75	5	0,5	-1	2,25	-7	1	-2	0,75	-8,25	-1	-5,5

B

	Aldh1L1	Arpp21	Cdh13	Cldn4	Cyp2C6v1	Esr1	Extl1	Gadd45	Gsn	Niacr1	Pten	Sirt4	Wap	Wif1
CC-LL	3	0,5	-7	0,5	10	-1	-4	0	0	-1	1,5	-4	6	3
CC-MC	5,2	1,75	-1,5	-0,5	1	-1	2,75	-3,25	-2,5	0	-2,25	-5,25	-4	-1
CC-MH	7,5	3	-6,5	-6	0	0	-3	-6	-0,5	-4,75	-2,5	-6	-6	-3
CC-CHC	4	5	-0,5	2,5	2	2	20,5	0	-0,5	8,5	-4	1,5	1	-2
CH-CHC	3	0	1,5	3	2	0	8	-1,5	-1	-4,5	-3,5	0	-10	2
CHC-LL	-1	-4,5	-6,5	-2	8	-3	-24,5	0	0,5	-9,5	5,5	-5,5	5	5
CHC-MC	1,25	-3,25	-1	-3	-1	-3	-17,7	-3,25	-2	-8,5	1,75	-6,75	-5	1
CHC-MH	3,5	-2	-6	-8,5	-2	-2	-23,5	-6	0	-13,2	1,5	-7,5	-7	-1
CH-CC	1	5	-2	-0,5	0	2	12,5	1,5	0,5	13	-0,5	1,5	11	-4
CH-LL	2	-4,5	-5	1	10	-3	-16,5	-1,5	-0,5	-14	2	-5,5	-5	7
CH-MC	4,25	-3,25	0,5	0	1	-3	-9,75	-4,75	-3	-13	-1,75	-6,75	-15	3
CH-MH	6,5	-2	-4,5	-5,5	0	-2	-15,5	-7,5	-1	-17,7	-2	-7,5	-17	1
LL-MC	2,25	1,25	5,5	-1	-9	0	6,75	-3,25	-2,5	1	-3,7	-1,25	-10	-4
LL-MH	4,5	2,5	0,5	-6,5	-10	1	1	-6	-0,5	-3,75	-4	-2	-12	-6
MH-MC	-2,25	-1,25	5	5,5	1	-1	5,75	2,75	-2	4,75	0,25	0,75	2	2
HH-CC	-6	-3	0,5	1,5	1	-1	-2	-0,5	-0,5	-2	5,5	1,5	0	0
HH-CH	-5	2	-1,5	1	1	1	10,5	1	0	11	5	3	11	-4
HH-CHC	-2	2	0	4	3	1	18,5	-0,5	-1	6,5	1,5	3	1	-2
HH-LL	-3	-2,5	-6,5	2	11	-2	-6	-0,5	-0,5	-3	7	-2,5	6	3
HH-MC	-0,75	-1,25	-1	1	2	-2	0,75	-3,75	-3	-2	3,25	-3,75	-4	-1
HH-MH	1,5	0	-6	-4,5	1	-1	-5	-6,5	-1	-6,75	3	-4,5	-6	-3

Supplemental figure 2: Statistical overview of 14 candidate genes for DNA methylation in Wistar rats.

For statistical analysis contrast tests based on the beta-regression model were used to make pairwise group comparisons which were adjusted for multiple testing (Tukey test). Different shades of grey indicate the negative decadal logarithm of three distinct significance levels (from light to dark: $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, white = not significant). Numbers in boxes correspond to median difference in methylation between corresponding groups. With (B) and without (A) E2 treatment. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

0 3.0

A

	Aldh1L1	Arpp21	Cdh13	Cldn4	Cyp2c6v1	Esr1	Extl1	Gadd45b	Gsn	Niacr1	Pten	Sirt4	Wap	Wif1
CC-LL	0.04	-0.13	2.86	-0.13	n.d.	-0.07	0.00	0.00	0.19	0.17	0.53	0.01	-0.05	-0.02
CHC-CC	-0.03	-0.05	-1.87	-0.14	n.d.	0.01	-0.04	0.00	12.04	0.00	0.11	-0.01	-0.75	0.01
CHC-CH	-0.02	-0.18	1.29	0.06	n.d.	1.34	-0.04	-0.01	0.74	0.24	-0.18	-0.02	0.29	0.02
CHC-LL	0.01	-0.18	0.99	-0.27	n.d.	-0.06	-0.04	0.00	12.23	0.17	0.64	0.00	-0.79	-0.01
CH-CC	-0.01	0.12	-3.17	-0.19	n.d.	-1.33	0.00	0.01	11.30	-0.24	0.29	0.01	-1.03	-0.01
CH-LL	0.03	0.00	-0.30	-0.32	n.d.	-1.40	0.00	0.01	11.49	-0.08	0.82	0.02	-1.08	-0.02
HH-CC	-0.02	0.06	-0.48	0.33	n.d.	0.23	0.02	0.00	4.31	-0.05	-0.01	0.01	0.63	0.03
HH-CH	-0.01	-0.06	2.69	0.53	n.d.	1.56	0.02	-0.01	-6.99	0.19	-0.29	0.00	1.67	0.04
HH-CHC	0.01	0.11	1.40	0.47	n.d.	0.22	0.06	0.00	-7.73	-0.05	-0.11	0.02	1.38	0.02
HH-LL	0.03	-0.07	2.39	0.20	n.d.	0.16	0.02	0.00	4.50	0.12	0.53	0.02	0.59	0.01

B

	Aldh1L1	Arpp21	Cdh13	Cldn4	Cyp2c6v1	Esr1	Extl1	Gadd45b	Gsn	Niacr1	Pten	Sirt4	Wap	Wif1
CC-LL	-0.06	-0.29	2.20	-0.16	n.d.	0.03	-0.02	-0.01	-1.34	-0.07	0.12	-0.02	0.10	-0.02
CHC-CC	0.06	-0.07	-2.34	-0.34	n.d.	-0.04	-0.06	-0.02	4.59	-0.12	0.24	0.01	-0.22	-0.07
CHC-CH	-0.01	-0.34	-3.14	-0.29	n.d.	0.51	-0.02	-0.03	1.24	-0.21	-0.16	0.01	-0.72	-0.06
CHC-LL	0.00	-0.36	-0.14	-0.50	n.d.	-0.01	-0.08	-0.02	3.26	-0.19	0.36	-0.01	-0.12	-0.09
CH-CC	0.07	0.26	0.80	-0.06	n.d.	-0.55	-0.04	0.01	3.35	0.10	0.40	0.00	0.50	-0.01
CH-LL	0.01	-0.02	3.01	-0.22	n.d.	-0.53	-0.06	0.00	2.02	0.03	0.52	-0.02	0.60	-0.03
HH-CC	0.06	0.23	1.35	-0.13	n.d.	-0.08	-0.05	-0.01	1.56	0.06	0.01	0.02	0.09	-0.01
HH-CH	-0.01	-0.03	0.55	-0.08	n.d.	0.47	-0.01	-0.02	-1.79	-0.03	-0.38	0.01	-0.41	0.00
HH-CHC	0.00	0.30	3.69	0.21	n.d.	-0.04	0.01	0.01	-3.03	0.18	-0.23	0.01	0.32	0.06
HH-LL	0.00	-0.06	3.56	-0.29	n.d.	-0.05	-0.07	-0.02	0.23	-0.01	0.14	-0.01	0.20	-0.03

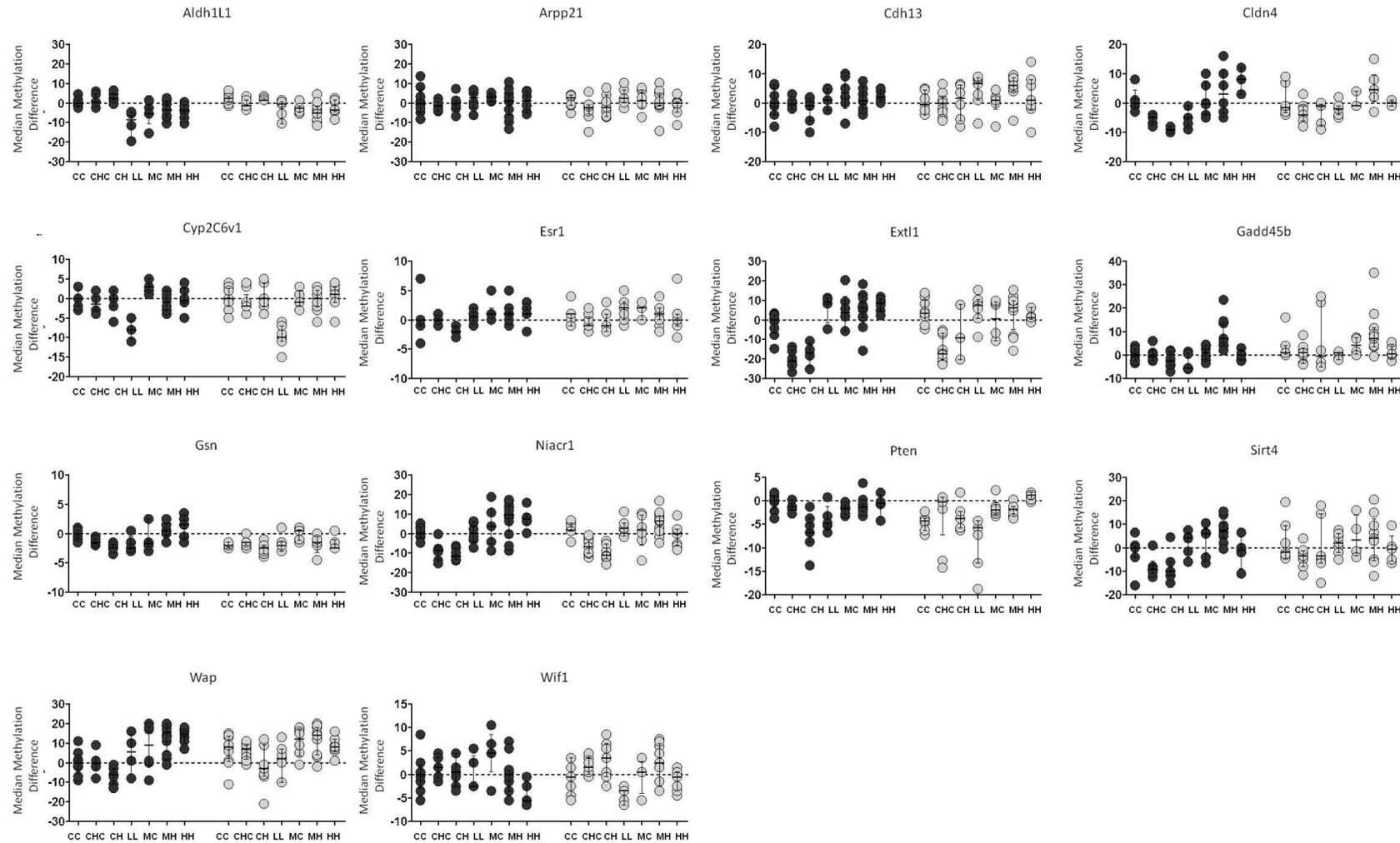
Supplemental figure 3: Statistical overview of 14 candidate genes for gene expression in Wistar rats.

For statistical analysis contrast tests based on the linear regression model were used to make pairwise group comparisons which were adjusted for multiple testing (Tukey test). Different shades of grey indicate the negative decadal logarithm of three distinct significance levels (from light to dark: $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, white = not significant). Numbers in boxes correspond to fold change expression compared to corresponding group. With (B) and without (A) E2 treatment. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

	Aldh1L1	Arpp21	Cdh13	Cldn4	Cyp2c6v1	Esr1	Extl1	Gadd45b	Gsn	Niacr1	Pten	Sirt4	Wap	Wif1
CH-CC											M			
CHC-CC									M					
MC-CC									M		M			
HH-CC							M			M	M			
HH-CH				M										
HH-CHC				M			M			M			M	
HH-LL				M			G							M

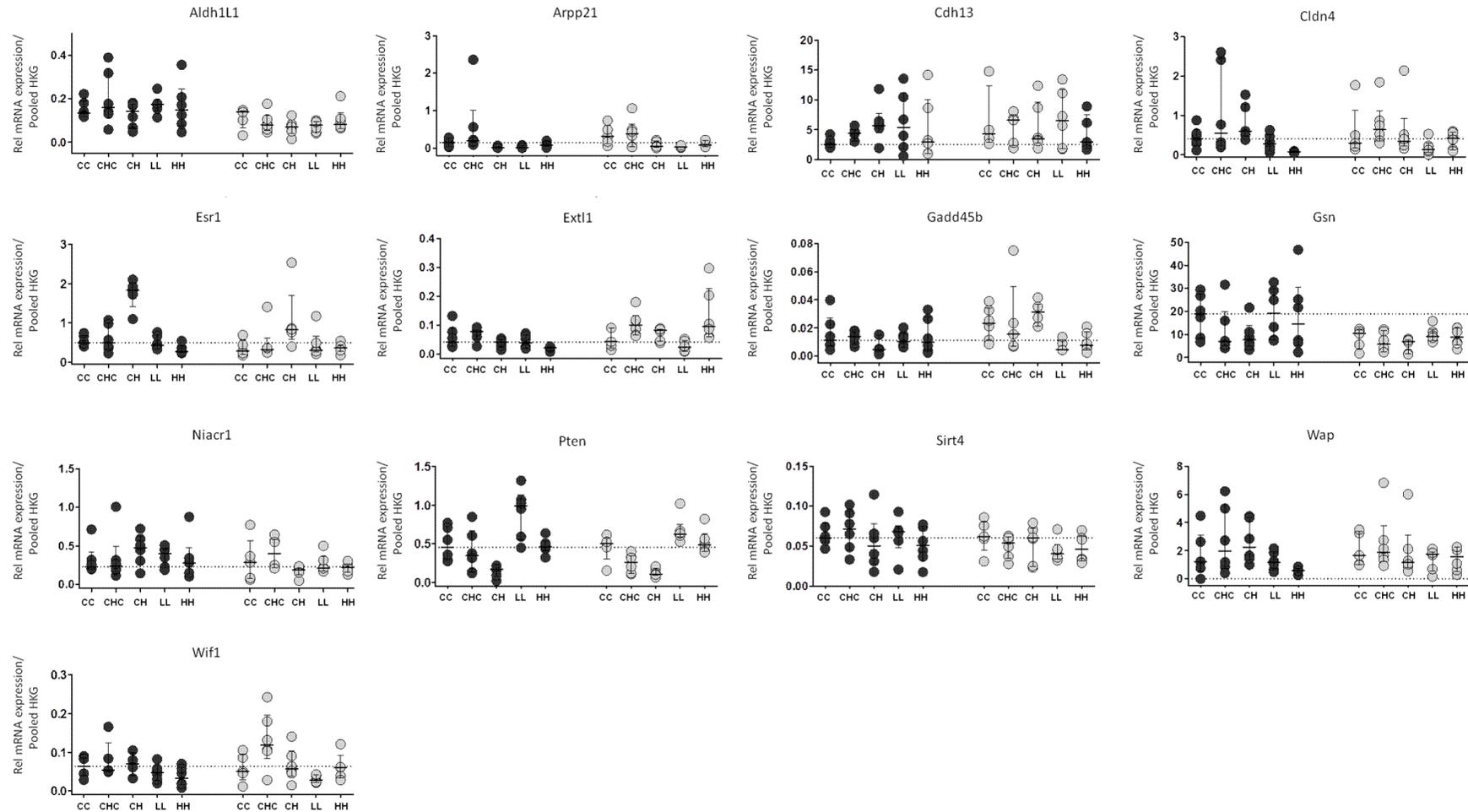
Supplemental figure 4: Statistical overview of interaction effects in 14 candidate genes in Wistar rats.

For statistical analysis interaction term was tested to identify E2-dependent group effects. Different shades of grey indicate the negative decadal logarithm of three distinct significance levels (from light to dark: $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, white = not significant). Letters in boxes indicate changes in median difference in methylation (M) or relative mRNA expression (G). Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).



Supplemental figure 5: Summary of DNA methylation across all experimental groups of Wistar rats.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Median methylation differences per group is indicated by a horizontal line, vertical lines indicate interquartile range. For group wise comparison see Supplemental figure 2. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).



Supplemental figure 6: Summary of relative mRNA expression across all experimental groups of Wistar rats.

Each dot in the aligned dot plot represents one animal (dark grey, without E2; light grey, with E2 stimulation from PND 94-97). Relative mRNA expression to pooled housekeeping genes (HKG) of the CC group is indicated as horizontal lines, with vertical lines depicting interquartile range. Median rel. expression of the CC group is indicated as horizontal dashed lines. For group wise comparison see Supplemental figure 3. Group names indicate dietary IF exposure relative to pre and post OVX (IF-free control diet (C), low (L), medium (M), and high (H) IF diets).

Publications and Poster Presentations

Publications in peer reviewed journals

Istas, G., Declerck, K., Pudenz, M., Szarc vel Szic, K., Heyninck, K., Haegeman, G., Gerhauser, C., Heiss, C., Rodriguez-Mateos, A., Vanden Berghe, W. (2016) Distinctive blood DNA methylation associated with atherosclerosis: identification of concordant DNA methylation changes of BRCA1 and CRISP2 in surrogate blood and carotid plaque material. (in preparation)

Gerhauser, C., Heilmann, K., and Pudenz, M. (2015) Genome-Wide DNA Methylation Profiling in Dietary Intervention Studies: a User's Perspective. **Current Pharmacology Reports** 1, 31-45.

Blei, T.*, Soukup, S.T.*, Schmalbach, K.*, Pudenz, M.*, Moller, F.J., Egert, B., Wortz, N., Kurrat, A., Muller, D., Vollmer, G., Gerhauser, C., Lehmann, L., Kulling, S. E., Diel, P. (2015) Dose-dependent effects of isoflavone exposure during early lifetime on the rat mammary gland: Studies on estrogen sensitivity, isoflavone metabolism, and DNA methylation. **Molecular Nutrition and Food Research** 59, 270-283 (*shared first authorship)

Pudenz, M., Roth, K., and Gerhauser, C. (2014). Impact of Soy Isoflavones on the Epigenome in Cancer Prevention. **Nutrients** 6, 4218-4272.

Poster Presentations

Isoflavones reduce estrogen-induced genome-wide DNA methylation changes in rat mammary tissue
Pudenz, M., Weichenhan, D., Lipka, D., Blei, T., Diel, P., Plass, C., Gerhäuser, C., *PhD poster presentation*, DKFZ, Heidelberg, Germany, 11/2013

Isoflavones reduce estrogen-induced genome-wide DNA methylation changes in rat mammary tissue
Pudenz, M., Weichenhan, D., Lipka, D., Blei, T., Diel, P., Plass, C., Gerhäuser, C., *Cell Symposia: Cancer Epigenomics*, Sitges, Spain 10/2013

DNA Methylation in Rat Mammary Tissue and Modulation by Isoflavones - Genome-wide sequence analysis for non-bioinformaticians

Pudenz, M., Weichenhan, D., Lipka, D., Blei, T., Diel, P., Plass, C., Gerhäuser, C., *PhD retreat*, Weil-der Stadt, Germany, 06/2013

Acknowledgements

First of all, I would like to thank Dr. Clarissa Gerhäuser for giving me the opportunity to transfer my nutrition scientist knowledge into the world of epigenetics. Thank you Clarissa, for giving me the chance to do my Ph.D. in your group Cancer Chemoprevention and Epigenomics and for all the advice and support you gave me throughout my studies. Thanks for all the thesis corrections, report approvals, active discussions and infinite ideas (also for the famous raspberry cake and the delicious Easter Bunny as well as Santa Claus surprises).

I also thank Prof. Christoph Plass for his scientific input and the opportunity to work in the Division of Epigenomics and Cancer Risk Factors at the DKFZ.

A special thanks goes to PD Dr. Odilia Popanda and Prof. Dr. Günter Vollmer for agreeing to be part of my thesis advisory as well as defense committee. Thank you for being available for discussions, scientific advices and constructive input. Further, my gratitude goes to Prof. Thomas Rausch and Dr. Michael Milsom, in advance, for acting as members of my defense committee as well as for reviewing and grading this thesis.

It is a pleasure to thank my collaboration partners, especially:

- Our partners from the IsoCross Project: Prof. Dr. Dr. Patrick Diel, Profs. Dr. Sabine Kulling and Dr. Achim Bub and Prof. Dr. Leane Lehmann. Thank you for many interesting discussions, telephone conferences and project meetings. Further, I would like to thank Dr. Frank Möller, Dr. Sebastian Soukup, Dr. Katja Schmalbach, Anne Kurrat and Dennis Müller for fantastic collaborative relations. A special thanks goes to Dr. Tina Blei who supported me far beyond her time in the project. Thanks Tina for your motivation, the "yes, you can" in replay mode and for being there whenever needed.

- Paul Datlinger and Dr. Christoph Bock from the CEMM in Vienna for helping me with the establishment of the RRBS library preparation protocol. Thank you Paul for all the skype calls, emails and endless advices.

- Prof. Michael Muders and Stephanie Zeiler (TA) from the Institute of Pathology at the TU Dresden for their help preparing the LCM samples and for labeling the DCIS on the sections.

Bioinformatics data analysis would not have been possible without the support of Dr. Daniel Lipka, Dr. Yassen Assenov and Dr. Lei Gu. Thanks for being always patient with me and for giving me elaborate advise when I was lost in the black box made of algorithms, scripts and the command line interface.

I would like to thank Dr. Thomas Hielscher for his support with the biostatistics analyses. Thank you Thomas, for your interest in the project, for your instant availability and for all the brainpower you put into our analysis.

Thank you Dr. Felix Bestvater and Manuela Brom for your kind introduction into the LCM technology and for the many 'Cha' tastings as well as diverting conversations during the last years.

For his scientific input particularly regarding the genome-wide techniques I would like to thank PD Dr. Dieter Weichenhan. Thanks Dieter for your interest during the Monday seminars, for being approachable whenever needed and for imprinting a word that I will always connect with you...

A big thank you goes to my office mate Katharina Heilmann. We shared the daily routine, the ups and downs, the success and the failure. Without you Kathi, lab time would have been much harder. Thank you, for the scientific and non scientific-related interest and your great willingness to help.

Thanks Cornelia Siebenkäs, Kevin Roth, Anna-Lena Krug and Tanja Maihöfer for help with different aspects of this work, for the recovering coffee and tea breaks, and for just having a good time. I look forward to the Marathon and more Gin and Tonic tastings! I also would like to thank my group mates Clarissa Feuerstein, Melanie Weiss and David Brocks for their scientific input in our group meetings as well as Janine Jung and Klara Klein for their excellent work during their practical.

To our technical support team Annette Weninger, Peter Waas, Reinhard Gliniorz, Marion Bähr, Monika Helf and Oliver Mücke, thanks for your great technical assistance and organizational efforts. A special thank you goes to Karin Klimo, my second office mate, who supported me with any lab-related issues. Thanks Karin for a your dedication, for your energy and of course for your legendarily delicious Alfredo pizza.

Further I would like to thank Susanna Grenner for her excellent help in administrative affairs.

All the members of C010 division who I might have missed:

Previous and current post-docs: Dr. David Scherf, Dr. Soo-Zin Kim-Wanner, Dr. Tania Witte, Dr. Annika Baude, Dr. Marina Laplana, Dr. Olga Bogatyrova, Dr. Khelifa Arab, Dr. Simin Öz, Dr. Reka Toth, and Dr. Michael Daskalakis.

My fellow Ph.D. students: Justyna Wierzbinska, Christoph Weigel, Christian Faltus, Mridul Nair and Sina Stäble.

Thank you for the helpful comments, numerous chats and crazy ideas you shared with me. It was great to get to know you, to work with you and to have one or the other relaxing BBQ at the Neckarwiese.

Meiner Familie danke ich besonders für ihre Unterstützung während dieser Zeit, für ihr Fragen und das ebenso wichtige nicht-Nachfragen, für ihr Vertrauen sowie für das Wissen ihrer bedingungslosen Liebe. Ohne sie wäre dies alles nicht möglich.

To my ladies, Marianne und Alexandra, danke, dass ich euch immer einen Satz warme Ohren verpassen durfte und ihr mir trotz der Endlosschleifen geduldig zugehört habt.

Zuletzt sei mein Dank an Stephan gerichtet, für das Rücken freihalten, für das dicke Fell, mit dem du meine Launen ertragen hast und für die Geduld, das die Zeit der Entbehrungen bald vorüber geht. Danke für dein Verständnis und für das Gefühl zu Hause zu sein.

Danke.