Transcriptional regulation of YAP1 and IL33 by CDX2 in colon epithelial cells

Larsen, Sylvester

Publication date:
2018

Document Version
Peer reviewed version

Transcriptional regulation of *YAP1* and *IL33* by CDX2 in colonic epithelial cells

PhD dissertation by

Sylvester Larsen

August 2018

This thesis has been submitted to the Doctoral School of Natural Science, Department of Science and Environment, Roskilde University

Transcriptional regulation of *YAP1* and *IL33* by CDX2 in colonic epithelial cells

PhD thesis – August 2018

Sylvester Larsen

Academic supervisors

Professor Jesper T. Troelsen, Department of Science and Environment, Roskilde University.

Associate Professor Ole B. V. Pedersen, Department of Clinical Immunology, Næstved Hospital.

Preface

This thesis is submitted to fulfil the partial requirements for obtaining a PhD degree at Roskilde University. The work presented has been carried out at the Department of Science and Environment at Roskilde University, Denmark and at the Department of Clinical Immunology, Næstved hospital, Denmark from January 2014 to August 2018, including leave of absence. The PhD dissertation is divided into 5 sections; an introduction to the field of research and the aim of the thesis, a results section containing three manuscripts, a discussion of the thesis work and aim, a conclusion, and future perspectives. Abbreviations and terminology are listed at the beginning of the thesis. The thesis results are based on two submitted and one published manuscripts. Paper I and II are the main contributions to the scope of the Ph.D. dissertation, and paper III is a multi-center collaboration effort, where collaborators contributed a large part of the acquired data.

The thesis work was supported by grants from the Independent Research Fund Denmark (grant ref: 4004-00140B) and the region Zealand foundation for health and research.

The two main academic supervisors on the project were:
Professor Jesper T. Troelsen, Department of Science and Environment, Roskilde University.
Associate professor Ole B. V. Pedersen, Department of Clinical Immunology, Næstved Hospital.

Acknowledgements

The completion of this thesis would not have been possible without the assistance and support of many people.

First of all, I would like to thank my supervisors Professor Jesper T. Troelsen and Associate Professor Ole B. V. Pedersen for their generous guidance, advice, contribution to my work, and not least their patience. Without their supervision, it would not have been possible to complete this work.

Secondly, I would like to extend my gratitude to Professor Claus Henrik Nielsen for our collaboration and my training together with his group at the Institute for Inflammation Research, Center for Rheumatology and Spine Diseases at Copenhagen University Hospital.

I would also like to thank all the co-authors on my research papers, especially Mehmet Coskun for his invaluable input to the IL-33 paper, and Dr Eric P. Bennett and his research group at the University of Copenhagen for their collaboration on the PrIITE paper.

A special thank you should be granted to my colleagues at the University of Roskilde and Næstved Hospital. All of you have contributed to a great working environment. Thanks to the Troelsen research group for the great collaboration and friendly atmosphere, both in and out of the lab. It has been a joy to work with both former and current colleagues of the group, and I am grateful for the time there. A special thanks go out to those I have worked most closely with during my employment: Johanne, Maiken, Morten, Steffen, Isabella, Anja, Julian, Dennis, and Michelle. Also thanks to the technicians that have given very helpful technical support, lab assistance and a friendly work environment.

ACKNOWLEDGEMENTS

I would also like to thank the gastro group members for the weekly meetings where we exchanged both great scientific ideas and small talk over cake. A special thanks go out to Professor Louise T. Dalgaard for many years of collaboration and scientific feedback.

I am very thankful for my family, which have given me support and encouragement throughout my Ph.D. work. It would have been very difficult to finish this work without them believing in me.

Last but in no way least, I would like to thank my wonderful girlfriend Biljana for her inspiration and invaluable support. Without her emotional backup, comforting assurances, and especially her overwhelming patience, this thesis would most likely not have been finished.

Summary

Colorectal cancer (CRC) and inflammatory bowel disease (IBD) are two major health problems that are increasing worldwide. More than 7400 people die from CRC in the Nordic countries alone, and more than 1 in 350 people are afflicted with IBD in the western part of the world. Dysregulation of colonic homeostasis has been linked to the development and severity of both diseases. A large number of genes and several molecular signalling pathways are involved in maintaining the homeostasis. A network of transcription factors controls the expression of the colonic genes. Caudal type homeobox 2 (CDX2) is an intestinal-specific transcription factor that plays a central role in this network, and it regulates many genes important for homeostasis. The dysregulation of CDX2 has been associated with the development and severity of CRC and IBD.

The goal of the thesis is to increase our knowledge about CDX2s role in colonic homeostasis, by identifying CDX2 target genes important for the colonic epithelium. Two potential CDX2 target genes, *YAP1* and *IL3*3 were identified by analyzing ChIP-Seq data from Caco-2 and LS174T colon epithelial cell lines. Both genes are expressed in many tissues and their role in most tissues are well described. However, their function in the colonic epithelium is not well understood. Investigating their connection with the intestinal specific transcription factor CDX2 will enable a deeper understanding of their role in colonic epithelium.

In paper I of this thesis, CDX2 was shown to transcriptionally regulate *YAP1* promoter activity in Caco-2 cells, partly through an identified intragenic enhancer. By using a CDX2 inducible colon epithelial cell line described in paper III, it was possible to demonstrate that the YAP1 protein level was increased by CDX2 induction. In paper II, *IL33* was found to be a target gene for CDX2 in LS174T colon epithelial cells. Furthermore, it was found that CDX2 was indispensable for *IL33* promoter activity

in LS174T cells, and that induction of CDX2 in the colonic epithelial cell model caused a dose-dependent increase in the *IL33* mRNA level.

Paper III describes a novel method for developing an inducible CDX2 knockout LS174T based cell line. This cell line expresses no endogenous CDX2 protein, but CDX2 expression can be induced using doxycycline. In addition, paper I and II demonstrated how the identification of novel CDX2 target genes is possible using the $LS174T^{CDX2-/-}$ cell model.

In summary, *YAP1* and *IL33* were identified as novel CDX2 target genes with importance for colonic homeostasis. It is hypothesized, that the regulation by CDX2 contributes to a *YAP1* and *IL33* intestinal-specific expression pattern that define their role in the colon homeostasis. Furthermore, it is likely that other transcription factors in the colon play a role in the transcriptional regulation of *YAP1* and *IL33*.

Danish summary

Kolorektal cancer (CRC) og inflammatorisk tarmsygdom (IBD) er to tiltagende globale sundhedsproblemer. Flere end 7400 mennesker dør af CRC i de nordiske lande alene, og i den vestlige verden er mere end 1 ud af 350 mennesker diagnosticeret med IBD. Misregulering af homeostasen i kolon er påvist at have betydning for både udviklingen og sværhedsgraden af sygdommene. Et stort antal gener og flere molekylære signalveje er involveret i at opretholde homeostasen i kolon. Ekspressionen af generne i kolon styres af et netværk af transkriptionsfaktorer. Den tarmspecifikke transkriptionsfaktor Caudal type homeobox 2 (CDX2) spiller en central rolle i dette netværk, og den regulerer mange gener der er vigtige for homeostasen. Dette kan blandt andet ses ved at både udviklingen og sværhedsgraden af CRC og IBD påvirkes når reguleringen af CDX2 er unormal.

Formålet med denne afhandling er at øge den samlede viden om CDX2s rolle i homeostasen i kolon ved at identificere gener som spiller en rolle i kolonvævet og som reguleres af CDX2. Gennem en analyse af ChIP-Seq data fra kolonepitelcellelinjerne Caco-2 og LS174T, blev de to gener *YAP1* og *IL33* identificeret som værende potentielle CDX2 regulerede gener. Begge gener er udtrykt i mange væv, og deres rolle i de fleste af disse væv er godt beskrevet. Dog er deres rolle i kolonepitelet stadig uklar. Ved at undersøge hvordan *IL33*, *YAP1* og CDX2 er forbundet, er det muligt at opnå en dybere forståelse for deres roller i kolonepitelet.

Manuskript I i afhandlingen viste at CDX2 regulerer *YAP1* promotoraktiviteten i Caco-2 celler gennem en hidtil ubeskrevet enhancer i *YAP1* genet. Herudover viste resultaterne fra CDX2-inducering i en LS174T baseret kolonepitelcellelinje, beskrevet i manuskript III, at mængden af YAP1-protein forøges i cellerne ved CDX2-induktion.

Resultaterne i manuskript II viste at CDX2 kan regulere *IL33* promoteraktiviteten i LS174T kolonepitelceller. Desuden blev det konstateret, at *IL33*

promotoraktiviteten var afhængig af tilstedeværelsen af CDX2 protein i cellerne og at CDX2-inducering forårsagede en dosisafhængig stigning i *IL33* mRNA niveauet.

Manuskript III beskriver en ny metode til udvikling af en LS174T baseret cellelinje. I cellelinjen produceres der intet CDX2 protein medmindre protein ekspressionen induceres ved stimulering med doxycylin. Derudover viste manuskript I og II hvordan det er muligt, at benytte LS174T$^{CDX2-/-}$ cellemodellen til at identificere hidtil ukendte gener som reguleres af CDX2.

Sammenlagt viser resultaterne fra manuskripterne at CDX2 regulerer både *YAP1* og *IL33*, som begge har betydning for homeostasen i kolonepitelceller. Dette giver anledning til en hypotese om at CDX2s regulering af YAP1 og IL33 tilfører dem et intestinal-specifikt ekspressionsmønster, som er med til at definere deres rolle i homeostasen i kolon. Herudover er det sandsynligt at andre transskriptionsfaktorer i kolon også spiller en rolle i den transskriptionelle regulering af *YAP1* og *IL33*.

Abbreviations

Gene names are written in italic, while proteins are in non-italic letters. Human protein symbols are in all caps, while mouse proteins have only the first letter capitalized. An example of annotation: Human protein: CDX2, Human gene: *CDX2*, mouse gene: Cdx2. Gene names from referenced materials might have been updated to their new annotations (Official symbols and official full gene names) in the thesis text, according to the guidelines for human gene nomenclature (Wain et al., 2002).

APC	APC, WNT signalling pathway regulator
ATOH1	Atonal bHLH transcription factor 1
AXIN2	Axin 2
BMP	Bone morphogenetic protein
CBC	Crypt base columnar cells
CD	Crohn's disease
CDH17	Cadherin 17
CDHR5	Cadherin related family member 5
CDX1	Caudal type homeobox 1
CDX2	Caudal type homeobox 2
CLDN1-4	Claudin 1-4
CRC	Colorectal cancer
CTCF	CCCTC-binding factor
EGF	Epidermal growth factor
ELMO3	Engulfment and cell motility 3
EP300	E1A binding protein p300
FCGBP	Fc fragment of IgG binding protein
FPKM	Fragments Per Kilobase Million
GATA4	GATA binding protein 4
GATA6	GATA binding protein 6
GSK3B	Glycogen synthase kinase 3 beta
GUCY2C	Guanylate cyclase 2C
H3K4me1	Methylation of Lys4 of histone H3

HEPH	Hephaestin
HNF1A	Hepatocyte nuclear factor 1 alpha
HNF4A	Hepatocyte nuclear factor 4 alpha
IBD	Inflammatory bowel disease
IEC-6	Intestinal cell line
IL-33	Interleukin 33
LAMC2	Laminin subunit gamma 2
LCT	Lactase
LGR5	Leucine rich repeat containing G protein-coupled receptor 5
MAPK	Mitogen-activated protein kinase
MEP1A	Meprin A subunit alpha
MUC2	Mucin 2
NHEJ	Non-homologous end joining
PRKDC	Protein kinase, DNA-activated, catalytic polypeptide
PTGS2	Prostaglandin-endoperoxide synthase 2
RPKM	Reads Per Kilobase Million
TALENs	Transcription-activator-like effector nucleases
TCF7L2	Transcription factor 7 like 2
TEAD	Transcription factor TEA domain family members
TNF	Tumour necrosis factor
TPM	Transcripts Per Kilobase Million
UC	Ulcerative colitis
YAP1	Yes-associated protein 1
ZFNs	Zinc-finger nucleases
ZG16	Zymogen granule protein 16

Vocabulary

ChIP-Seq	Chromatin immunoprecipitated DNA
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/CRISPR-associated 9
DNase-seq	DNase sequencing
Hippo pathway	Signalling pathway in animals that controls organ size through various regulatory networks
RNA-Seq	RNA expression profiling with microarray or sequencing
Wnt pathway	Signalling pathway that regulates genes important for cell fate determination

List of publications

Publications included in the thesis

- **Paper I** – HNF4α and CDX2 regulate *YAP1* promoter activity in intestinal cell lines

  Sylvester Larsen[1,2], Johanne Davidsen[1,3], Katja Dahlgaard[1], Ole B. Pedersen[2], and Jesper T. Troelsen[1*]

  [1]Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark; [2]Department of Clinical Immunology, Næstved Hospital, Ringstedgade 77B, 4700 Næstved, Denmark. [3]Department of Surgery, Center for Surgical Science, Zealand University Hospital, Lykkebækvej 1, 4600, Køge, Denmark

  Manuscript is submitted to the Journal of Cellular Biochemistry.

- **Paper II** – CDX2 regulates interleukin-33 gene expression in intestinal epithelial cell lines.

  Sylvester Larsen[1,2], Jakob Benedict Seidelin[3,] Johanne Davidsen[1,4], Katja Dahlgaard[1], Claus Henrik Nielsen[5], Eric Paul Bennett[6], Ole Birger Pedersen[2], Mehmet Coskun[7,3#], and Jesper Thorvald Troelsen[1*#]

  [1]Department of Science and Environment, Roskilde University, Roskilde, Denmark. [2]Department of Clinical Immunology, Næstved Hospital, Næstved, Denmark. [3]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Herlev, Denmark. [4]Department of Surgical Gastroenterology, Zealand University Hospital, Køge, Denmark. [5]Institute for Inflammation Research, Center for Rheumatology and Spine Diseases, section 7521, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark. [6]Copenhagen Center for Glycomics, Department of Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [7]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark.

  Manuscript submitted to the journal Cellular and Molecular Gastroenterology and Hepatology

- **Paper III** – Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression
  Rita Pinto[1,2,3,4], Lars Hansen[4], John Hintze[4], Raquel Almeida[1,2,3,5], **Sylvester Larsen**[6,7], Mehmet Coskun[8,9], Johanne Davidsen[6], Cathy Mitchelmore[6], Leonor David[1,2,3], Jesper Thorvald Troelsen[6] and Eric Paul Bennett[4,*]
  [1]i3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [2]Ipatimup, Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal, [3]Faculty of Medicine of the University of Porto, Porto, Portugal, [4]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark, [5]Department of Biology, Faculty of Sciences of the University of Porto, Porto, Portugal, [6]Department of Science and Environment, Roskilde University, Roskilde, Denmark, [7]Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark, [8]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark and [9]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark.

  Nucleic Acids Research 2017 Jul 27;45(13):e123. doi: 10.1093/nar/gkx371.

List of additional published papers

1 **"Intestinal regulation of suppression of tumorigenicity 14 (ST14) and serine peptidase inhibitor, Kunitz type -1 (SPINT1) by transcription factor CDX2"** Scientific Reports, doi:10.1038/s41598-018-30216-z, Danielsen ET, Olsen AK, Coskun M, Nonboe AW, **Larsen S**, Dahlgaard K, Bennett EP, Mitchelmore C, Vogel LK, Troelsen JT. **(2018)**

2 "**The VTI1A-TCF4 colon cancer fusion protein is a dominant negative regulator of Wnt signaling and is transcriptionally regulated by intestinal homeodomain factor CDX2**" PLOS ONE, doi:10.1371/journal.pone.0200215, Davidsen J, Larsen S, Coskun M, Gögenur I, Dahlgaard K, et al. **(2018)**

3 "**Peptoids successfully inhibit the growth of gram negative *E.coli* causing substantial membrane damage"** Scientific Reports, doi:10.1038/srep42332,

Biljana Mojsoska, Gustavo Carretero, **Sylvester Larsen**, Ramona Valentina Mateiu and Håvard Jenssen, (**2017)**

4 **"Cellular inhibitor of apoptosis protein 2 controls human colonic epithelial restitution, migration, and Rac1 activation"** American Journal of Physiology: Gastrointestinal and Liver Physiology, doi:10.1152/ajpgi.00089.2014, Seidelin, J. B., **Larsen, S**., Linnemann, D., Vainer, B., Coskun, M., Troelsen, J. T., & Nielsen, O. H. **(2015)**

5 **"MicroRNA-29a is up-regulated in beta-cells by glucose and decreases glucose-stimulated insulin secretion"** Biochemical and Biophysical Research Communications, doi:10.1016/j.bbrc.2012.08.082, Bagge A, Clausen TR, **Larsen S**, Ladefoged M, Rosenstierne MW, Larsen L, Vang O, Nielsen JH, Dalgaard LT. **(2012)**

Contents

LIST OF PUBLICATIONS

Aim of the PhD project

Diseases of the colon are a major and increasing health problem worldwide. In the Nordic countries, colorectal cancer (CRC) kills more than 7400 people every year and more than 1 out of 350 people live with inflammatory bowel disease (IBD) in most of the Western world. Dysregulation of colonic homeostasis has been linked to the disease development and severity of both IBD and CRC. The colonic homeostasis is affected by many factors, such as diet, commensal bacteria, and genetics, which influence the expression of colonic genes. A network of transcription factors and several major signalling pathways controls gene expression in the colon. The transcription factor CDX2 plays a central role in this network. Dysregulation of CDX2 and its target genes have been linked to problems in homeostasis, and to IBD and CRC. To increase our understanding of the complex pattern of transcriptional regulation in the colon, it is necessary to identify the target genes of CDX2 and their function thereby enabling a deeper understanding of these diseases.

Overall aim

The aim of this thesis was to investigate the role of CDX2 in the expression of colonic genes with importance for the homeostasis and disease development in the colon. Molecular methods and model systems aimed to identify CDX2 target genes in colon cells was employed and development.

Specific aims of the research are:

- *Investigating whether the expression of YAP1 is regulated by CDX2 in colon cells and describing the mechanism of regulation*
- *Investigating whether expression of IL-33 is regulated by CDX2 in colon cells, and describing the mechanism of regulation*
- *Developing a cell model that can be used to investigate the effects of varying CDX2 levels on target gene expression in the colon.*

1 Introduction

1.1 When colon homeostasis fails - typical pathologies of the colon

When intestinal homeostasis fails to be maintained properly, due to environmental or inherited factors, pathologies such as unregulated growth, dysfunctional or abnormal cells, or inflammatory diseases can occur.

CRC is the third most frequent cancer worldwide with an estimated incidence rate of 1.36 million cases recorded in 2012. Additionally, it is one of the most deadly types of cancer, killing almost 700.000 people yearly (M. Ervik, 2016). Based on the statistics on colorectal cancer by the World Cancer Research Fund International, it is predicted that the worldwide number of cases by 2035 will rise extensively. The statistics on the newly diagnosed cancers in 2012 for the top world cancers can be seen in (Figure 1A). In Europe in 2012, CRC is estimated to take up 32.9 % of the total number of incident cancer and 31% of the total number of cancer deaths worldwide (Figure 1B, mortality rate not shown). In Denmark, the CRC incidence rate is the second highest of the Nordic countries (Engholm G, 2018).

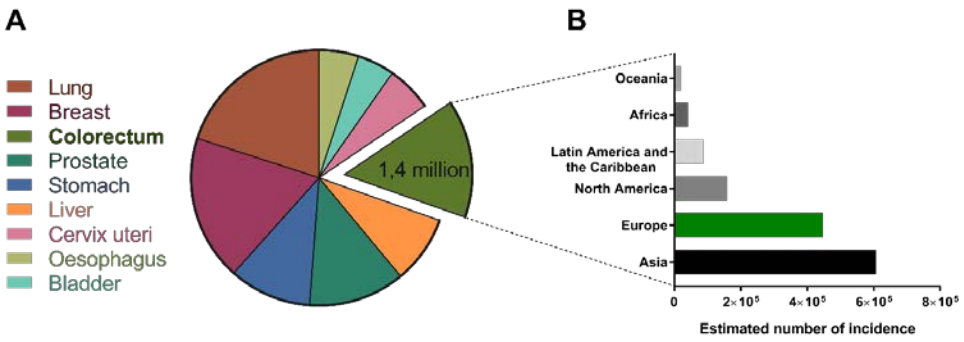


Figure 1 - World cancer statistics. A) Data shows number (1000s) of newly diagnosed cancers for both male and female population in 2012 for the most common cancers (M. Ervik, 2016). B) Estimated number of incident cases of CRC worldwide in 2012 (both male and female) (M. Ervik, 2016).

The incidence of CRC is age-dependent, where getting CRC before age of 50 is rare for both men and women (Haggar & Boushey, 2009). In Denmark, the risk of getting CRC before age 75 is 5.4 % and 4.1 %, for men and women, respectively (Engholm G, 2018). Risk factors for developing colorectal cancer include both environmental and hereditary origins. About 20-30 percent of colorectal cancers are caused by hereditary conditions, leaving the rest of the cases to undetermined origins such as environmental factors like diet (Haggar & Boushey, 2009).

Dysregulated homeostasis can also result in the development of inflammatory bowel disease (IBD), which is an umbrella term that contains the two major inflammatory diseases ulcerative colitis (UC) and Crohn's disease (CD), and an array of less common diseases such as microscopic colitis. UC is only present in the colon, while CD also affects the small intestine.

The incidence and prevalence rates of IBD are not as well-documented as for cancer, but a 2018 systematic review compared 147 studies and found that the prevalence of inflammatory bowel disease exceeded 0.3% in North America, Oceania, and many countries in Europe, with lower rates in other continents. The data shows that the incidence of inflammatory bowel disease in North America and Europe is stabilizing,

but that it is rising in newly industrialized countries in Africa, Asia, and South America with an annual percentage increase of up to 11% for some countries (Ng et al., 2018).

1.2 Homeostasis in the colon

One of the key organs in the human body, for sustaining healthy digestion and regulating water and salt balance needed for maintaining homeostasis is the large intestine. In humans, the large intestine begins at the cecum, continues into the colon, rectum, and ends at the anal canal. The colon is the longest part of the large intestine and it consists of four segments; the right-sided ascending colon, the transverse colon, the left-sided descending colon, ending in the sigmoid colon that is connected to the rectum. The colon is responsible for absorption of water, salts, and some nutrients like vitamins, excreting ions like potassium chloride and forming a protective cell-sheet that prevents harmful substances from underlying cells. The cells that line the colon form the colonic epithelium, which is organized into crypts of Lieberkühn, small invaginations of the mucosal surface. The crypts are considered the basic functional units of the colon, and its cells are responsible for maintaining colonic homeostasis and regeneration of the epithelium. H^3 thymidine labelling has shown that the entire colonic crypts, with the exception of a few stem cells, are renewed every 3-4 days (Lipkin, Bell, & Sherlock, 1963). Diseases of the colon such as inflammatory bowel diseases (IBD) and colon cancer are often linked to dysregulation of the processes responsible for cell renewal and differentiation in the crypts. The intestinal homeostasis is usually maintained by a constant renewal and differentiation of epithelial cells originating from pluripotent intestinal stem cells at the bottom of the crypts. The stem cells divide and differentiate into the four major cell types that make up the majority of colonic epithelium cells: colonocytes, goblet cells, enteroendocrine cells, and the deep crypt secretory cells (Humphries & Wright, 2008; Sasaki et al., 2016). Colonic crypts are similar to those of the small intestines, but they contain less diverse subpopulations of cells (H. C. Clevers

& Bevins, 2013). Furthermore, colon epithelium lacks the villus formations present in the small intestines (Sancho, Batlle, & Clevers, 2004).

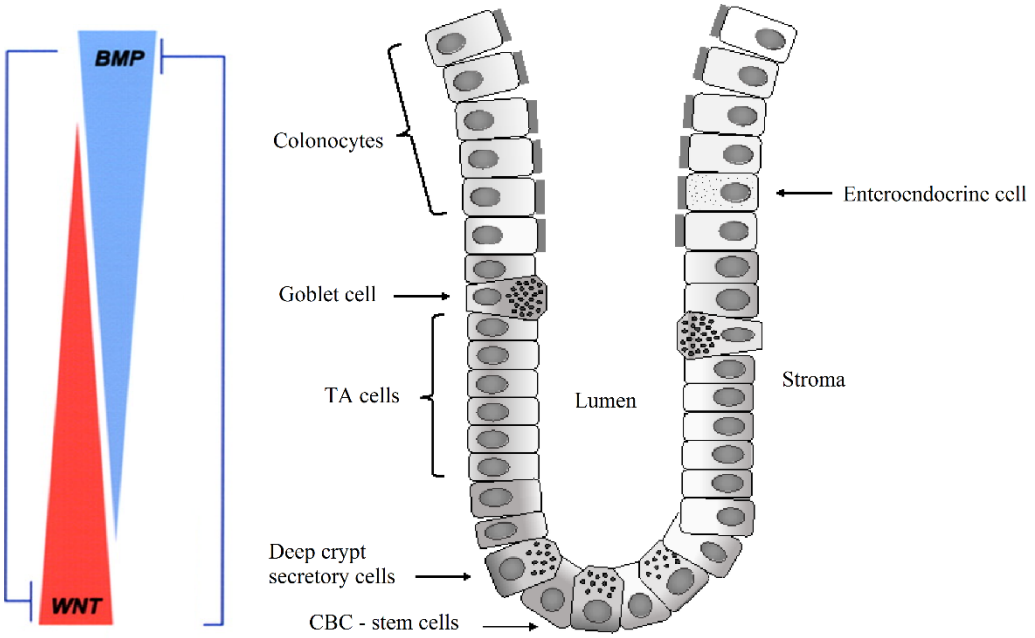


Figure 2 - Schematic representation of a colonic crypt and the level of WNT and BMP, Modified from (Kosinski et al., 2007; Testa, Pelosi, & Castelli, 2018).

To uphold homeostasis, the differentiated colonic cells must constantly be able to take up necessary nutrients and support the stem cells, while protecting the epithelium from luminal bacteria and regulating the hormonal balance of the gut. Colonocytes are the most abundant cell type in the colon, and their primary function is to absorb nutrients from the lumen through their microvilli surface and export them basally into the body. The colon supports a large microbial population which produces a significant amount of vitamin K and vitamins B_1, B_3, B_7, and B_9 (Thiamine, Niacin, Biotin, and Folate) often exceeding the amount available from dietary sources and it has been shown that colonocytes are able to take up these vitamins (Hill, 1997; Said, 2013). The vitamins are important for host nutrition but especially important for the colonocytes metabolism.

Goblet cells are responsible for producing the mucus layer of the colon. In the colon, a viscous inner mucus layer about 200-300 µm thick protects the epithelium from contact with the commensal bacteria, while a less dense outer layer serves natural habitat for the bacteria. The mucus consists of water and proteins called mucins with attached glycans that is used as a food source the commensal bacteria that inhabits the outer layer (Johansson, Larsson, & Hansson, 2011). The thickness of the mucus layer and the number of bacteria in the colon increases in a proximal-distal direction and under normal circumstances, it completely blocks direct contact between colonic bacteria and epithelial cells (Pelaseyed et al., 2014). The main function of the goblet cells is to secrete "mucin 2, oligomeric mucus/gel-forming" (MUC2), a glycoprotein that is used as a matrix for forming the mucus layer in cooperation with other mucus forming components (Pelaseyed et al., 2014). Some of these components are Fc fragment of IgG binding protein (FCGBP) which covalently binds and cross-links mucus proteins and zymogen granule protein 16 (ZG16) which can bind carbohydrates and affect cell proliferation (Johansson, Thomsson, & Hansson, 2009; Mito et al., 2018).

Enteroendocrine cells comprise only a small minority of about 0.4% of the overall epithelial cell population within the colonic crypts (Tsubouchi & Leblond, 1979) but still play a role in the gut physiology. Colonic enteroendocrine cells grow small extensions of the cell aimed towards adjacent epithelial cells and often have microvilli that extend towards the luminal surface. Common for enteroendocrine cells is their production of hormones contained in secretory vesicles that can be exocytosed in response to membrane depolarization caused by neuronal signals (Gunawardene, Corfe, & Staton, 2011). Colonic enteroendocrine cells appear in three distinct subtypes, the enterochromaffin, L, or D cells. The enterochromaffin cell type are the most abundant subtype with about 70% in the proximal colon, falling to 40% in the rectum, while the absolute number of enteroendocrine cells are almost constant throughout the colon (Cristina, Lehy, Zeitoun, & Dufougeray, 1978; Sjolund, Sanden, Hakanson, & Sundler, 1983). Enterochromaffin cells secrete 5-HT (Serotonin) that regulates intestinal motility

and secretion, and appetite. L cells are the second most abundant enteroendocrine cell in the colon, and their main secretory products are Peptide YY, that suppress appetite and stimulates colonocytes differentiation, and proglucagon-derived peptides like GLP-1 and GLP-2 that induces the incretin effect or stimulates colonocytes proliferation, respectively (Gunawardene et al., 2011). Very few D cells are found in the colon, where they mainly secrete somatostatin that inhibits digestive hormone release and stimulates peristalsis.

The deep crypt secretory cells are located at the bottom of the colonic crypts, intermingled with crypt base columnar cells (Figure 2). Both *in vivo* and *ex vivo* colonic organoid experiments revealed that the crypt base columnar cells survival depends on the presence of the deep crypt secretory cells (Sasaki et al., 2016). Furthermore, it was found that the deep crypt secretory cells express signal molecules like Notch ligands and epidermal growth factor similar to the function of the Paneth cells found in the small intestines, although they do not express Wnt3 ligand as Paneth cells do. The function of the deep crypt secretory cells seems to correlate well with that of Paneth cells in the small intestines (Sasaki et al., 2016). In some cases when homeostasis is dysregulated due to inflammatory diseases or cancers of the colon, regular small intestine Paneth cells have been found in the ascending colon tissue (Humphries & Wright, 2008).

The self-renewal of the colonic epithelium is carried out by the intestinal stem cells residing at the bottom of the colonic crypts. In 2007, it was established that pluripotent stem cells could be identified using the marker leucine rich repeat containing G protein-coupled receptor 5 (LGR5) which have later been validated as a bona fide stem cell marker (Barker & Clevers, 2010; Barker et al., 2007). In the colon the stem cells are the crypt base columnar cells (CBC) which divide about once a day, producing progenitor cells referred to as transit-amplifying cells that rapidly divide and differentiate into the absorptive or secretory cells of the colon epithelium (Biswas et al., 2015). The colon lacks the second intestinal stem cell population present in the small intestines,

called the +4 label-retaining cells, which are more quiescent and functions as secretory cell precursors and as a backup cell pool for damaged tissue (Buczacki et al., 2013).

The homeostasis of the colonic epithelium renewal is controlled by the signals in the stem cell niche which is the microenvironment surrounding the crypts. The niche controls cell expansion and proliferation by the fine-tuning of morphogenic signal molecules along the crypt. Expression analyses of genes in cells from colonic crypts and bottoms involved in signal pathways have revealed significant differences in expression of genes in the bone morphogenetic protein (BMP), Notch, and Wnt, signalling pathways (Figure 2) (Kosinski et al., 2007). It has been shown that the Notch pathway is required to repress the transcription factor atonal bHLH transcription factor 1 (ATOH1), leading progenitors to abandon differentiation to the secretory cell fate. The Wnt/β-catenin signalling has also been connected to the cell fate decision between absorptive and secretory cells by its interaction with Notch signalling in ATOH1 regulation, by active Wnt blocking its ubiquitination and degradation, reviewed in (Noah, Donahue, & Shroyer, 2011). The TGFβ superfamily contains the BMP signalling, which phosphorylates SMAD proteins for signal transduction. BMP represses Wnt signalling in the colonic stem cell niche, creating a gradient with high Wnt highest at the apical surface. This gradient is necessary for homeostasis in the stem cell niche, reviewed in (Biswas et al., 2015; Noah et al., 2011). The canonical Wnt/β-catenin signal pathways regulate colonic gene expression through enabling accumulation of β-catenin in the nucleus, which acts as a co-transcription factor or interacts directly with other regulators (H. Clevers & Nusse, 2012).

1.3 Gene regulation is essential to maintain homeostasis in the colon

Maintaining homeostasis in the colon depends on the accurate control of cell proliferation and differentiation, through an integration of both external and internal signal pathways. The ability to control cell signalling comes from the cells ability to control the expression of their gene through gene regulation. As all cells in the human body generally share the same genome, the transcriptional molecular machinery is utilized to control, modify and differentiate gene expression to fit the needs of specific cell types or tissues. Transcriptional regulation is the first and one of the most important steps in controlling the amount of gene product produced in a cell. Controlling gene expression at the transcriptional level is the mechanism by which cells can limit or block any transcripts from being produced, saving the cell from expending energy on subsequent post-transcriptional regulatory mechanisms.

1.3.1 Intestinal gene regulation is controlled by a network of intestinal transcription factors

Several transcription factors play a major role in regulating gene expression in the intestine. A network of intestinal transcription factors controls the expression of many genes important for proper function of the intestine. A key component of the network is Caudal type homeobox 2 (CDX2), and it also includes other major transcription factors, such as Caudal type homeobox 1 (CDX1), hepatocyte nuclear factor 1 alpha and 4 alpha (HNF1A and HNF4A), transcription factor 7 like 2 (TCF7L2), and GATA binding protein 4 and 6 (GATA4 and GATA6) (Boyd, Bressendorff, Møller, Olsen, & Troelsen, 2009; Gao, White, & Kaestner, 2009; Olsen, Boyd, Danielsen, & Troelsen, 2012). Some of these transcription factors are not intestinal-specific, but combined and in conjunction with intestinal-specific TF's such as CDX2, intestinal-specific regulation emerges from they collected unique pattern of regulation. Understanding the role of each transcription

factor in the network can help gain knowledge of the molecular mechanisms of homeostasis dysregulation. Due to the complexity of the regulatory networks, we are still far from a complete understanding of the roles of the intestinal transcription factors (Figure 3).

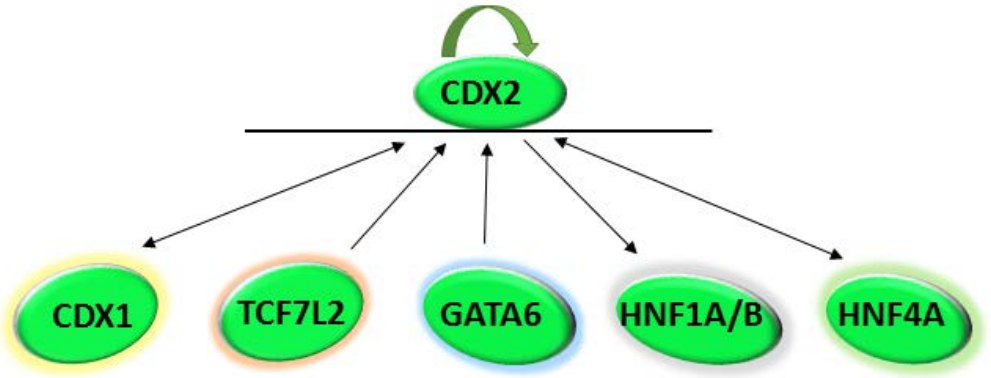


Figure 3 – Schematic illustration of the intestinal transcriptional network. CDX2 auto-regulates its own expression, and the expression of CDX1, HNF1A/B, and HNF4. CDX1, TCF7L2, GATA6, and HNF4A regulate CDX2 expression.

Animal and cell models have revealed that CDX2 expression is regulated by GATA6, TCF7L2, and HNF4A synergistically in colon cancer cells (Benahmed et al., 2008; Boyd et al., 2009). Additionally, it was shown that CDX2 auto-regulates its own expression through binding to its promoter and to a distal enhancer near the 3' end of the gene (Barros et al., 2011; Boyd et al., 2010; Xu, 1999). It has also been shown that CDX2 is able to regulate the transcription factors HNF4A, HNF1 alpha and beta (HNF1A/HNF1B), and CDX1 which are major transcription factors for genes in the intestines and shown to be important for expression of intestine-specific genes and enterocyte differentiation (Boyd et al., 2010). A clear cooperation has been observed between CDX2, HNF4A, and GATA4 to maintain intestinal homeostasis by assembling

at tissue-specific *cis*-regulatory sites. Deficiency in crypt cell replication have been shown to occur as a result of CDX2 and GATA4 inactivation in intestinal cells, whereas inactivating CDX2 and HNF4A impaired maturation and viability of colonocytes (San Roman, Aronson, Krasinski, Shivdasani, & Verzi, 2015). HNF4A is a key transcription factor important for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium (Cattin et al., 2009). HNF4A is highly expressed in the intestines, but contrary to CDX2, it is also expressed in other tissues (Drewes, Senkel, Holewa, & Ryffel, 1996). HNF4A is essential for the formation of crypts in the colon and loss of HNF4A alters the secretory cell lineage in the colon by increasing goblet cell abundance (Cattin et al., 2009; Garrison et al., 2006). Furthermore, it was shown to have a role in gluconeogenesis in the colon cells, by inducing the expression of the key enzyme glucose-6-phosphatase (Gautier-Stein, Zitoun, Lalli, Mithieux, & Rajas, 2006).

In general, transcription factors can be divided into two main groups, the general transcription factors, and sequence-specific transcription factors. The general transcription factors make up the basal transcriptional complex, of which the main enzyme is RNA polymerase II. RNA polymerase II and its required general transcription factors, control DNA unwinding, initiation of transcription, and RNA transcript elongation (Tupler, Perini, & Green, 2001). Even though the basal transcriptional machinery is sufficient to enable gene transcription, the gene-specific control that maintains intestinal homeostasis are exerted by the sequence-specific transcription factors through their binding to regulatory elements such as promoters, enhancers, silencers and insulators, reviewed in (Chatterjee & Ahituv, 2017; Naar, Lemon, & Tjian, 2001).

Sequence-specific transcription factors are divided into distinct families defined by the type of DNA domains they contain, and several major families have been identified and characterized. The two largest families are the zinc finger proteins and the homeodomain transcription factors (Tupler et al., 2001). One characteristic of the zinc finger family is their dependence on zinc-atoms, which when bound mediate the

conformation of the proteins needed for the binding to the DNA. Zinc finger proteins recognize relatively long (20-40 bp) DNA sequences compared to most other transcription factors (Fedotova, Bonchuk, Mogila, & Georgiev, 2017). Homeodomains consist of approximately 60 amino acids which are folded into a structure closely related to a helix-turn-helix motif often found in prokaryotes (Mitchell & Tjian, 1989). They vary significantly in sequence composition within the family but still exhibit nearly identical DNA docking (Noyes et al., 2008). Some transcription factors contain several types of binding domains, increasing the number of sequences they can bind to on the DNA. Furthermore, there is a great similarity in the sequence recognition of different transcription factor domains, which enables them to bind to the same DNA sequences (Tupler et al., 2001).

Some intestinal sequence-specific TFs are able to affect gene regulation indirectly by changing the chromatin structure. Remodelling of the chromatin structure is important for eukaryotic gene regulation because it can limit the access of both sequence-specific transcription factors and the basal transcriptional machinery to the DNA. The structure can either be as condensed and transcriptionally inert heterochromatin or as open and transcriptionally active euchromatin, depending on the post-translational modified nucleosomes (Reddy, Park, & Natarajan, 2012). The histone code is the pattern of modifications that include hyper and hypo-methylation and acetylation as well as phosphorylation, ubiquitination, and sumoylation (Jenuwein & Allis, 2001; Saksouk, Simboeck, & Dejardin, 2015).

Research into gene regulation by sequence-specific transcription factors is often performed by investigating the direct DNA binding and subsequent effect of the transcription factor on attracting the basal transcriptional machinery. However, general co-activators that modify chromatin availability also exerts significant regulatory activity. One example of this is the indirect effect of CDX2's and HNF4A's corporation with E1A binding protein p300 (EP300), a histone acetyltransferase that regulates basal gene transcription via chromatin remodelling. EP300 binding sites have been identified

as overlapping with both CDX2 and HNF4α binding sites in ChIP-Seq experiments (Boyd et al., 2010; Rosenbloom et al., 2013; Verzi et al., 2010). This co-localization might suggest that EP300 is able to bind to CDX2, which would demonstrate another layer of complexity to the regulation of intestinal gene expression (Hussain & Habener, 1999). Furthermore, the chromatin structure can be affected by the simultaneous binding of multiple TFs in the same region. In eukaryotic cells, the DNA-binding domains of transcription factors typically recognize relatively short sequences that are clustered closely together on the genome (Wunderlich & Mirny, 2009). Many of these sequences contain multiple TF binding sites in stretches of DNA with an open chromatin structure often found in promoters and enhancers. It has been shown that the association of TFs at promoters or enhancers stabilizes the DNA-protein interactions of other TFs in the vicinity (Panne, 2008). With these many types of regulatory mechanisms in mind, it becomes clear that we need a very detailed understanding of the transcription factors interactions to make predictions and interventions that affect the general homeostasis of the colonic epithelium.

1.4 Biological role of proteins and genes investigated in this thesis

The following sections will highlight the role and the characteristics of the proteins that are investigated in this thesis. The main protein of interest for our investigation of the colonic homeostasis is CDX2 due to its key role. Additionally, the transcriptional co-activator YAP1 and the potential transcription factor IL-33 were studied.

1.4.1 CDX2

CDX2 is an intestinal-specific transcription factor expressed in the small intestine and the colon epithelium of adult mammals (James & Kazenwadel, 1991; Silberg, Swain, Suh, & Traber, 2000). CDX2 is a 313 amino acid homeobox transcription factor containing a TAD (Rings et al., 2001) and a homeodomain for DNA binding (Figure 4).

Figure 4 – Structure of isoform 1 (NP_001256.3) of the human CDX2 protein. TAD: Transactivation domain; HD: homeodomain. Numbers represent the position of functional domains in the 313 aa long protein.

CDX2, which is a key part of the intestinal transcription factor network, is a sequence-specific transcription factor, which binds to distinct sequences on the DNA, and modulates the ability of the general transcription factors to bind and initiate mRNA transcription. Sequence-specific transcription factors mediate regulation through interactions between DNA binding domain(s) to the regulatory sequences of the DNA while TAD(s) or effector domain(s) mediate binding to other proteins (Mitchell & Tjian, 1989). The DNA binding domain gives the transcription factor its specificity while the TAD is essential to the TFs ability to activate or repress gene expression by indirect or direct interaction with other transcription factors (Mitchell & Tjian, 1989; Naar et al., 2001). Sequence-specific transcription factors are modular in their nature and can build large protein complexes able to fine-tune gene expression, through the use of the TAD (Frietze & Farnham, 2011). In general, the types of interactions that the TAD mediates can be divided into at least three groups: interaction with the basal transcriptional machinery and its co-factors, other sequence-specific TFs, or chromatin modifying enzymes, reviewed in (Frietze & Farnham, 2011).

In early embryonic stages, CDX2 is expressed together with two other homeobox transcription factors from the caudal family, CDX1, and CDX4, but throughout adult life, only CDX1 and CDX2 persist (Benahmed et al., 2008). An expression gradient is formed in the crypt-villus axis during development, with CDX2 primarily in the villus and CDX1 in the crypts (Silberg et al., 2000). In adults, CDX1 and CDX2 are expressed in the intestinal epithelium in all cells along the crypt-villus axis (James & Kazenwadel, 1991). The CDX2 expression is limited to the intestines,

partly by the utilization of promoter switching (Benahmed et al., 2008). Furthermore, CDX1 is believed to be fine-tuning the activity of CDX2 but was found to be dispensable for normal intestinal development as long as CDX2 is present (Bonhomme et al., 2008; Olsen et al., 2012).

1.4.1.1 The main role of CDX2

The importance of CDX2 in the intestinal tissue mainly lies in its ability to regulate expression of genes specific to the intestines or to modulate expression of widely expressed genes adding an intestinal regulatory dimension to the expression of such genes. CDX2 regulates several intestinal genes important for the intestinal function, including membrane-spanning transport proteins, enzymes involved in metabolism, and transcription factors (Figure 5). CDX2 also regulates a wide array of genes that are not intestinal-specific.

An essential function of the intestine is the transport and breakdown of nutrients from the lumen, often carried out by active transporters and enzymes. CDX2 regulates expression of guanylate cyclase 2C (*GUCY2C*), a key receptor for heat-stable enterotoxins and important for colon cell proliferation and tumorigenesis, and Cadherin 17 (*CDH17*) which mediates intestinal cell-cell adhesion (Basu, Saha, Khan, Ramachandra, & Visweswariah, 2014; Di Guglielmo, Park, Schulz, & Waldman, 2001; Hinoi et al., 2002). CDX2 also regulates meprin A subunit alpha (*MEP1A*) which is an enzyme that breaks down peptides and proteins and plays a role in inflammatory bowel disease and colon cancer, and hephaestin (*HEPH*) which main function is the transport of dietary iron from intestinal epithelial cells into the circulatory system (Boyd et al., 2010; Coskun et al., 2012; Hinoi et al., 2005).

CDX2 also regulates genes involved in digestive processes that break down nutrients so they can be taken up by intestinal cells. One example is the lactase (LCT) gene, which codes for an enzyme responsible for the breakdown of lactose to

simpler sugars that can be taken up by the intestinal cells (Mitchelmore, Troelsen, Spodsberg, Sjöström, & Norén, 2000; Troelsen et al., 1997).

The importance of CDX2 for the integrity and morphology of the intestinal lining is well defined since dysregulation of CDX2 leads to loss of the cells intestinal identity, failure of differentiation, and cell migration issues. These phenotypic traits might be a consequence of dysregulation of CDX2 target genes such as those that code for structural proteins. It was shown that CDX2 regulates cadherin related family member 5 (*CDHR5*) a Ca2+-dependent cell-cell adhesion molecule with an important role in brush border assembly in the intestinal epithelium (Hinkel et al., 2012). Other examples include Claudin 1-4 (*CLDN1-4*), that are integral membrane proteins found exclusively at tight junctions, or the extracellular matrix glycoprotein laminin subunit gamma 2 (*LAMC2*), that is important for cell migration and is considered a molecular marker of invading cancer cells (Bhat et al., 2012; Coskun et al., 2017). Another important protein involved in cell migration and morphology forming is engulfment and cell motility 3 (*ELMO3*) which might be important for villus formation in intestinal cells (Coskun, Boyd, Olsen, & Troelsen, 2010). Additionally, several genes in the Wnt/β-catenin degradation complex was found to be regulated by CDX2. The genes are APC, WNT signalling pathway regulator (*APC*), glycogen synthase kinase 3 beta (*GSK3B*), and axin 2 (*AXIN2*), all with essential functions in the intestinal epithelium (Coskun et al., 2014).

CDX2 can be regulated by post-transcriptional regulation, in the form of phosphorylation of the protein at specific amino acids. Mitogen-activated protein kinase (MAPK) mediates serine 60 phosphorylation of CDX2 in proliferating intestinal cells, leading to a decreased transactivation ability and spatial relocation of CDX2 whereas differentiating cells remain largely unphosphorylated (Rings et al., 2001). However, CDX2 phosphorylation by the MAPK family member p38 increases the transactivation ability of CDX2 during cell differentiation (Houde et al., 2001). A third example of

phosphorylation is cyclin-dependent kinase 2 mediated Ser-281 phosphorylation, which leads to poly-ubiquitination and degradation of CDX2 (Boulanger et al., 2005).

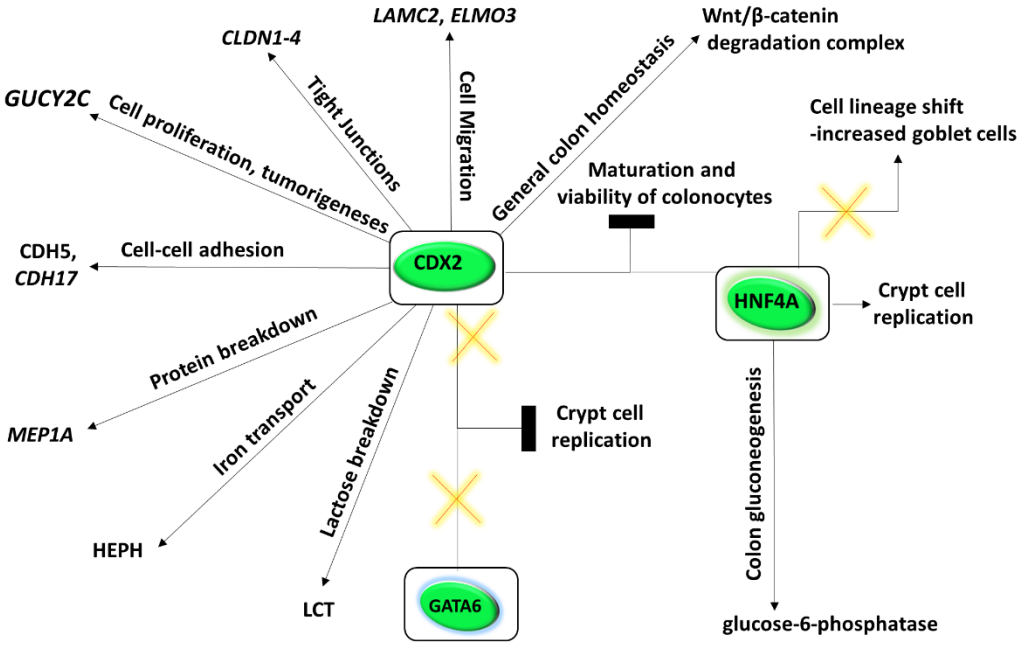


Figure 5 – Schematic illustration of the function of CDX2 in intestinal cells. Some functions are affected by CDX2 and cooperating transcription factor(s). Arrows point to genes that CDX2 regulates, and text along the arrows are the function affected by the particular gene. When a function is affected by a lack of a particular gene, an X blocks the line between them. A black box indicates an inhibition of the function.

1.4.1.2 Non-transcriptional actions of CDX2

Even though the main role of CDX2 is transcriptional regulation of genes important for intestinal actions, several studies have identified non-transcriptional mechanisms of CDX2 that are independent of DNA binding and transactivation. One study showed that CDX2 plays a role in the cell cycle by blocking ubiquitination of CDK inhibitor $p27^{Kip1}$, thus stabilizing it and inhibiting cell cycle progression. This effect had been evident even in mutated CDX2 DNA binding domains, showing that CDX2 also exerts non-

transcriptional actions (Aoki et al., 2011). This is in agreement with another study's findings that CDX2 with a mutated homeodomain was able to bind to β-catenin, a key component of the Wnt signalling pathway. This blocked β-catenin's association with the transcription factor TCF7L2 and inhibited Wnt signalling by blocking transcription (Guo, Funakoshi, Lee, Kong, & Lynch, 2010). Similarly, it was found that CDX2 intercepts and binds the P65 subunit of the NF-κB pathway in the nucleus, thus inhibiting the expression of prostaglandin-endoperoxide synthase 2 (PTGS2) which is overexpressed in colorectal cancers (S. P. Kim et al., 2004; Mutoh, Hayakawa, Sakamoto, & Sugano, 2007). Another study found that a splice-variant of CDX2 lacking the first helix in its DNA binding domain, was able to modulate the splicing patterns of genes expressed in the gut (Witek et al., 2014).

Furthermore, CDX2 was also shown to play a non-transcriptional role in DNA-repair. CDX2 lacking the transactivation domain was able to decrease the activity of protein kinase, DNA-activated, catalytic polypeptide (PRKDC) which is a kinase that facilitates the repair of DNA damage in colon cancer cells (Renouf et al., 2012). Interestingly, a study looking at colon cancer metastasis found that CDX2 expression was found to highly correlate with expression of mismatch repair enzymes and suggested that CDX2 might be involved in the loss of repair function in the cancer cells (Toth et al., 2018).

1.4.1.3 The role of CDX2 in development and maintenance of the intestinal epithelium

CDX2 is essential in the development and maintenance of intestinal tissue and loss of CDX2 leads to the epithelium losing its intestinal identity (Hryniuk, Grainger, Savory, & Lohnes, 2012; Silberg et al., 2000; Suh & Traber, 1996; Verzi et al., 2010). The role of CDX2 in embryonic development has been demonstrated by several studies using Cdx2 knockout mice. One study found that double allele Cdx2 knockout caused death

of embryos 3.5 days post coitum, which is the time that Cdx2 is normally expressed in the trophectoderm. Furthermore, in this study, it was demonstrated that heterozygous mutants exhibit stunted growth and develop intestinal cancers (Chawengsaksophak, James, Hammond, Köntgen, & Beck, 1997).

A study using conditional intestinal Cdx2 knockout mice showed that the lack of Cdx2 resulted in the dysregulation of pro-intestinal transcription factors Cdx1, Hnf1a, and Hnf4a, and caused intestinal epithelium to be replaced by keratinocytes (Gao et al., 2009). In an *in-vitro* intestinal organoid model, it was found that ablation of CDX2 leads to the development of gastric-like tissue lacking intestine-specific enzymes (Simmini et al., 2014). The importance of CDX2 in the initiation of cell differentiation is clearly seen in a study where CDX2 was overexpressed in the intestinal cell line (IEC-6), which caused an arrest of proliferation and initialization of differentiation (Suh & Traber, 1996). Furthermore, other *in vitro* studies have shown that CDX2 is important for determining the developmental fate of trophectoderm differentiation (Niwa et al., 2005; Ralston & Rossant, 2008).

1.4.1.4 Clinical relevance of CDX2

Apart from its role in controlling and maintaining intestinal development, CDX2 is also believed to act as a tumour suppressor in colorectal cancers. One study showed that mice with a decreased Cdx2 expression due to knockout of one allele of Cdx2 were more susceptible to develop malignant tumours in the colon then treated with a DNA mutagen (Bonhomme et al., 2003). In another study using heterogeneous knockout mice and intestinal cells lines, it was shown that lower CDX2 levels caused chromosomal aberrations and development of adenomatous polyps (Aoki, Tamai, Horiike, Oshima, & Taketo, 2003).

Immunostaining of colorectal adenocarcinomas from patients revealed that late-stage tumours were more likely to have lost CDX2 expression than early-stage tumours, thus revealing an inverse correlation between tumour progression and CDX2 expression. Furthermore, colorectal tumours with a low grade of differentiation showed a higher level of CDX2 than highly differentiated tumours (Bakaris, Cetinkaya, Ezberci, & Ekerbicer, 2008; Choi et al., 2006). Recently, CDX2 was identified as a prognostic biomarker in stage II and stage III colon cancer. The authors revealed that CDX2 negative tumours were correlated with a lower 5-year survival rate (Dalerba et al., 2016). In contrast to the evidence for CDX2 as a tumour suppressor, one study concluded that CDX2 is a lineage-survival oncogene in colorectal cancers and that data from the widely used mice models might not translate to humans (Salari et al., 2012).

CDX2 have been shown to interact with the well-established inflammatory signalling pathway Wnt and with the multifunctional proinflammatory cytokine "tumour necrosis factor" (TNF) (Coskun, 2014). The role of CDX2 in inflammation is still poorly understood, but studies involving mice and fruit flies have linked dysregulated CDX2 levels to increased intestinal permeability and pathogenic changes in the intestinal microbiome (Calon et al., 2007; Ryu et al., 2008). Several genes that are involved in the development of IBD are also regulated by CDX2, reviewed in (Coskun, 2014).

1.4.2 YAP1

Yes-associated protein 1 (YAP1), is a key transcriptional co-activator and a main downstream effector protein in the Hippo signalling pathway. The Hippo pathway is important for regulation of organ size, tissue homeostasis, and tumorigenesis (B. Zhao, Li, & Guan, 2010; Bin Zhao, Li, Lei, & Guan, 2010). This pathway regulates the switch between cell proliferation and differentiation (Pan, 2010; Bin Zhao et al., 2010). Both extra- and intracellular growth signals are relayed through the Hippo phosphorylation kinase cascade that terminates in the inactivation YAP1.

Of the eight different detected isoforms of YAP1, isoform 1 is commonly used in the literature and referred to as the consensus sequence. It is 504 amino acids and contains several conserved domains, like two WW domains that mediate protein-protein interactions, a transactivation domain, and a TEAD binding domain that can bind to the transcription factor TEAD4 (Chen et al., 1997; Li et al., 2010). These domains are responsible for the activity of YAP1, by regulating its shuttling to and from the nucleus (Attisano & Wrana, 2013; Imajo, Miyatake, Iimura, Miyamoto, & Nishida, 2012). The function of YAP1 is thus highly regulated through phosphorylation of specific amino acids. This can result in spatial control, but also inactivation or degradation of YAP1. However, the main role of YAP1 as a transcriptional co-activator is performed when it accesses the nucleus to regulate gene expression with other transcription factors. Transcriptional activators that do not bind DNA are called co-activators and are further divided into either the gene-specific or the "true" co-activators. YAP1 is a gene-specific co-activator. "True" co-activators regulate gene expression in a more general fashion like the chromatin remodeling done by EP300 (Ma, 2011).

YAP1 regulates the expression of many genes involved in proliferation, adhesion, migration and the extracellular matrix organization, mainly through its cooperation with the transcription factor TEA domain family members (TEAD) (Lian et al., 2010; Stein et al., 2015). *YAP1* is ubiquitously expressed and is found in a relatively medium expression level in the colon (Uhlen et al., 2015). As a key proliferation regulator, YAP1 has unsurprisingly been linked to cancer progression. Dysregulation of YAP1 is a significant factor in the development of colorectal cancers (Liang, Zhou, Zhang, Li, & Zhang, 2014; Steinhardt et al., 2008; Tschaharganeh et al., 2013). It has also been shown that overexpression of nuclear YAP1 in colorectal cancer tumours correlates with a poor prognosis in patients, but also that controlled YAP1 expression is necessary for colonic tissue repair and remodelling (Yui et al., 2018; Zygulska, Krzemieniecki, & Pierzchalski, 2017)

Not much is known about the transcriptional regulation of *YAP1* specifically in colon cells, but one study showed that the transcription factor SOX2 is an activator of *YAP1* expression in osteoprogenitor cells (Seo et al., 2013). Additionally, the transcription factors ERG, ETS transcription factor (ERG) and TEAD4 was demonstrated to increase *YAP1* expression through increasing acetylation of its promoter in prostate tumors (Nguyen et al., 2015).

1.4.3 IL-33

Interleukin (IL)-33 is a member of the IL-1 family of cytokines and is constitutively expressed in most cell types, including a relatively high expression in the intestines (Baekkevold et al., 2003; Moussion, Ortega, & Girard, 2008). It is synthesized as a full-length 30 kDa bioactive peptide and is typically located in the nucleus (Talabot-Ayer, Lamacchia, Gabay, & Palmer, 2009). IL-33 can be released from cells as an “alarmin” that amplifies immune responses following necrosis or damage caused by stress conditions such as injury, infections, and inflammation (Cayrol & Girard, 2014; Martin & Martin, 2016). Other cells can sense this alarm signal through a heterodimeric receptor complex comprising the co-receptor IL-1 receptor accessory protein (IL-1RAcP) and the transmembrane receptor suppressor of tumorigenicity 2 (ST2) and (Lingel et al., 2009) IL-33s regulation of gene expression is a less studied function. The full-length protein was revealed to bind to chromatin and negatively regulate transcription through a chromatin-binding motif (Carriere et al., 2007; Roussel, Erard, Cayrol, & Girard, 2008).

The combined cytokine and gene regulatory properties of IL-33 account for its important role for diseases of the colon. There is evidence that IL-33 has both a protective and pathogenic role in IBD, most likely by its interaction with the surrounding immune cells in the colon. A study on experimentally induced chronic colitis in mice has reported on a protective role of IL-33 during treatment, by its ability to decrease inflammation and shift Th1-directed cytokine responses toward a Th2-like pattern

(Grobeta, Doser, Falk, Obermeier, & Hofmann, 2012). Another DSS induced mice study has found that acute colitis was aggravated by the increase in pro-inflammatory cytokines in colon lymphocytes (Zhu et al., 2015). Several other studies report that IL-33 is involved in both acute and chronic colitis and colonic inflammation (Imaeda et al., 2011; Oboki et al., 2010; Pushparaj et al., 2013; Sedhom et al., 2013). Additionally, IL-33 was found upregulated in UC (Kobori et al., 2010; Seidelin et al., 2010). The role of IL-33 in IBD is still debated, for review see (Braun, Afonina, Mueller, & Beyaert, 2018). However, population-based cohort studies have found that patients suffering from IBD have an increased risk of developing colorectal carcinomas (Ekbom, Helmick, Zack, & Adami, 1990; Jess, Gamborg, Matzen, Munkholm, & Sorensen, 2005). Elevated expression levels of IL-33 has been detected in tumour tissues of CRC patients and in low-differentiated human CRC cells (Liu et al., 2014). Both protective and pathogenic roles of IL-33 have been suggested, but similar to the case with IBD, the role of IL-33 in CRC is still not very well described, for review see (Wasmer & Krebs, 2016).

1.5 Genetic modification as a tool for investigating gene expression

Genetic modification has been a useful tool for studying gene function in eukaryotic cells since the early 1970's. At first, gene modification was achieved by random mutagenesis, making researches able to investigate outcomes on phenotype and cell fate from gene abrogation caused by random mutations (Goth & Rajewsky, 1972). Through the 1980s a technique termed gene targeting was introduced, using homologous recombination to created designed genomic modifications at specific locations (Capecchi, 2005). This method relied on the ability of the cellular machinery to insert an artificially introduced DNA molecule into its genome by homologous recombination. The initial limitation of this method was the extremely low efficiency of homologous recombination in eukaryotic cells (J. S. Kim, 2016). In 1988, it was discovered that introducing double-strand breaks in the genome led to a significantly increased efficiency of homologous recombination (Rudin & Haber, 1988). This discovery led to one of the first demonstrations of highly efficient targeted gene modification in mammalian cells in 1994, where researchers expressed a meganuclease in mouse 3T3 cells containing its specific 18bp recognition domain. This introduced double-strand breaks at the recognition sites, which were subsequently repaired by the cells by either non-homologous end joining (NHEJ) or by homologous recombination when a donor template was present (Rouet, Smih, & Jasin, 1994). The NHEJ pathway can generate gene knockouts by inserting random repair sequence into break sites thus introducing nonsense mutations. Using meganucleases conferred a very specific recognition through their relatively long recognition sites, but they were not easily programmable and thus not suitable for many genome-editing purposes (Epinat et al., 2003).

At the beginning of the 1990s, researchers found a way to separate the DNA-binding domain and DNA-cutting nuclease domain from the restriction enzyme FoKI and showed that it was possible to combine these into novel sequence-specific endonucleases (Y. G. Kim & Chandrasegaran, 1994). A few years later zinc-finger

nucleases (ZFNs) were created by fusing the FokI nuclease domain to zinc-finger proteins, which are proteins containing at least three zinc-finger domains each interacting with up to 3 bp (Y. G. Kim, Cha, & Chandrasegaran, 1996). This makes them able to recognize relatively specific DNA sequences 18-36 bp long. ZFNs were used to modify human cells for the first time in 2003 and was widely used for several years despite some challenges with off-target effects and cytotoxicity (J. S. Kim, 2016; Porteus & Baltimore, 2003). ZFN's have been used to generate small deletions and insertions interrupting the reading frames of genes by taking advantage of the NHEJ repair pathway or to correct SNPs or insert whole genes using the HDR (Figure 6) (Smith, Deloukas, & Munroe, 2018).

The problem of cytotoxicity observed when using ZFNs was decreased by the introduction of a new type of chimeric proteins called transcription-activator-like effector nucleases (TALENs) in 2011 (Miller et al., 2011). TALENs are constructed by fusing several repeat domains of 33-35 amino acids from in the transcriptional activators of plant genes, each able to specifically bind one target base (Miller et al., 2011). Both ZFNs and TALENs have been extensively used for gene editing due to their high specificity and ability to be reprogrammed. However, constructing the nucleases for new targets is costly and heavily time-consuming, involving several cloning, purification, and testing steps for each target sequence. This is one of the reasons why they are starting to be replaced by a new gene editing method called the CRISPR/Cas9 system, short for **C**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats/CRISPR-associated 9. The CRISPR/Cas9 system is derived from bacteria, and in general the method consists of cutting DNA by incorporating a short guide RNA (gRNA) into the endonuclease Cas9 that subsequently guides Cas9 to the complementary DNA strand where it generates a double-stranded break (Gasiunas, Barrangou, Horvath, & Siksnys, 2012; Horvath & Barrangou, 2010; Wiedenheft, Sternberg, & Doudna, 2012). CRISPR/Cas9 has been utilized for eukaryote genome editing purposes in a variety of ways since 2013 (Cong et al., 2013). In recent years the method have seen significant improvements, such as the

development of modified or alternative endonucleases. Some of these are being designed to reduce off-target effects or to be able to activate or repress gene expression (Gilbert et al., 2013; Jinek et al., 2012; Maeder et al., 2013; Qi et al., 2013). Many types of modifications to endonucleases, gRNA sequences and the bioinformatics tools to detect targets and design gRNA are currently under development, for reviews see (Brooks & Gaj, 2018; Wilson, O'Brien, & Bauer, 2018; Wu, Lebbink, Kanaar, Geijsen, & van der Oost, 2018). Even with the advances of CRISPR based techniques, there is still great applicability of the established gene editing methods utilizing ZFNs and TALENs.

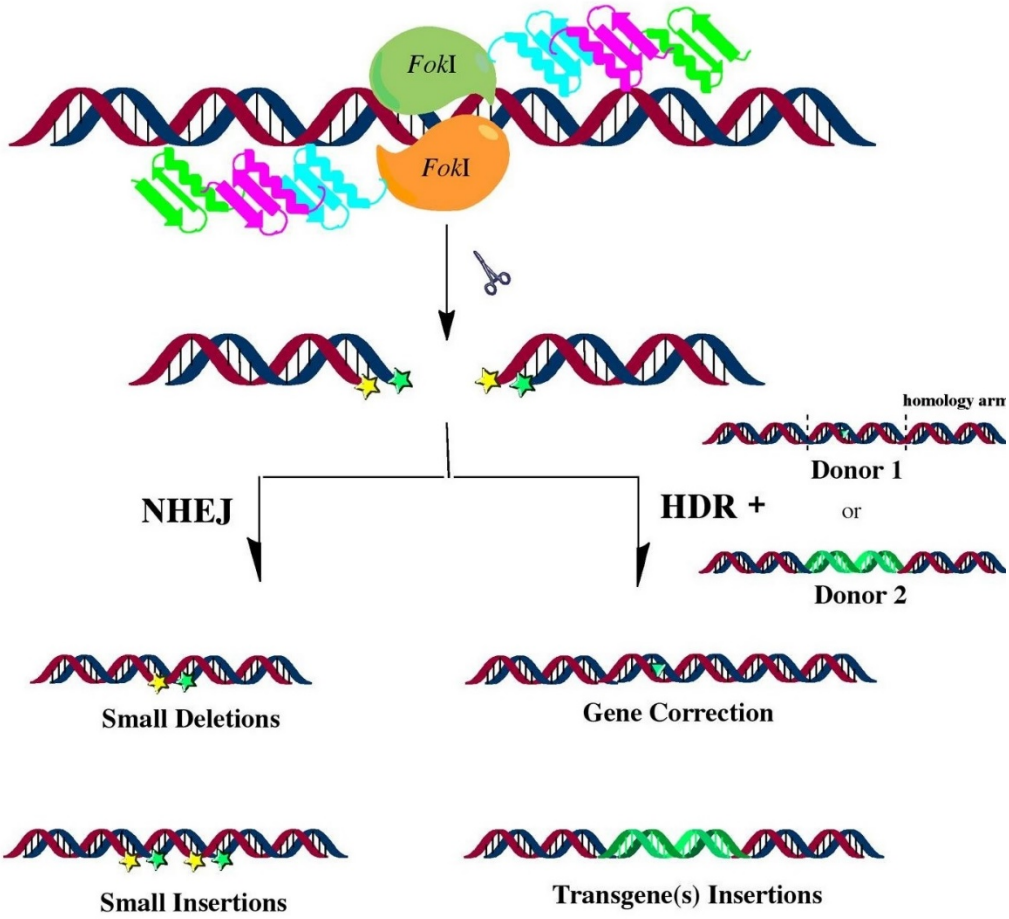


Figure 6 – Schematic representation of the Zinc-finger nuclease editing of genes and its applications. ZFNs bind to and cuts the DNA, that is then repaired by the cells by either Non-homologous End Joining (NHEJ) or Homologous-Directed Repair (HDR), modified from (Chou, Leng, & Mixson, 2012).

1.6 Identifying gene regulatory regions and target genes regulated by transcription factors

An array of methods have become crucial tools to identify and analyse the regulation and function of target genes in human cells. Usually, these include identification of gene regulatory elements or regions containing i.e. promoters and enhancers in the genome and determining what TFs binds to the regions. Often genome-wide scale assays, such

as DNase sequencing (DNase-seq), sequencing of chromatin immunoprecipitated DNA (ChIP-Seq), or RNA expression profiling with microarray or sequencing (RNA-Seq) are employed for this purpose.

1.6.1 RNA-Seq

RNA-Seq is a powerful tool to investigate the transcriptome, which is the complete set of RNA transcripts in a cell, and their quantity, at a specific time point (Figure 7). RNA-Seq enables high-throughput detection of gene expression on a whole genome level. The method consists of purifying RNA transcripts of interest, converting them to cDNA and attaching specific adapters, sequencing using next-generation sequencing methods, followed by bioinformatic analysis (Wang, Gerstein, & Snyder, 2009). Depending on the research question in focus, each of these steps can be performed vastly different. When searching for target genes for a particular transcription factor, like CDX2, the first step would be a purification of RNA that includes the mRNA from cells with and without CDX2. By determining the sequence of the transcripts, they can be mapped to the human genome and compared. Target genes can then be detected by searching for genes with different quantities of mapped transcripts in the two cell populations. The results of RNA-Seq experiments are often reported in Reads Per Kilobase Million (RPKM), Fragments Per Kilobase Million (FPKM) or Transcripts Per Kilobase Million (TPM). These metrics attempt to normalize for sequencing depth and gene length. To calculate RPKM first reads are normalized for sequencing depth by counting the total reads in a sample and dividing by 1 million, then this number is divided by the length of the gene in kilobases.

1.6.2 DNase-Seq

DNase-Seq identifies functional regulatory regions by detecting an open chromatin structure, but it is unable to ascertain whether the elements are promoters, enhancers, silencers, or insulators (Figure 7). In short, the assay works by exposing native DNA to DNase that will only digest accessible chromatin, followed by sequencing and mapping of the resulting DNA fragments by massive parallel sequencing and bioinformatics (Song & Crawford, 2010).

1.6.3 Chip-Seq

Chip-Seq is a more targeted approach than DNase-Seq that can be used to identify gene regulatory regions by the sequencing of DNA that has been co-precipitated with antibodies recognizing specific TFs (Figure 7). This approach can also be used to infer the function of the regulatory regions by detecting histone marks correlating to known functions. The first step in the ChIp-Seq method is cross-linking DNA-bound proteins to the DNA using formaldehyde, after which the DNA is extracted and fragmented. The fragments are then incubated with antibodies and the resulting immunoprecipitations are affinity purified and eluted. Cross-links are reverted while RNA and proteins are digested, and the resulting DNA purified and sequenced.

An example of an often-used ChIP-Seq analysis is the detection of active promoters by the identification of regions with increased methylation of Lys4 of histone H3 (H3K4me1) and H3K27Ac modifications simultaneously, whereas closed or inactive enhancers is identified by lack of H3K27Ac and high levels of H3K4me1 modifications (Bernstein et al., 2005; Creyghton et al., 2010; Rada-Iglesias et al., 2011). The transcriptional regulator “CCCTC-binding factor” (CTCF) is able to bind insulator elements, thus regions containing CTCF binding sites and/or H3K9me1 and H3K4me1/2/3 histone modifications indicate insulator elements that separate regulatory

regions (Barski et al., 2007). Silenced regions were strongly correlated with H3K27me2/3 and moderately correlated with H3K9me2/3 modification (Barski et al., 2007).

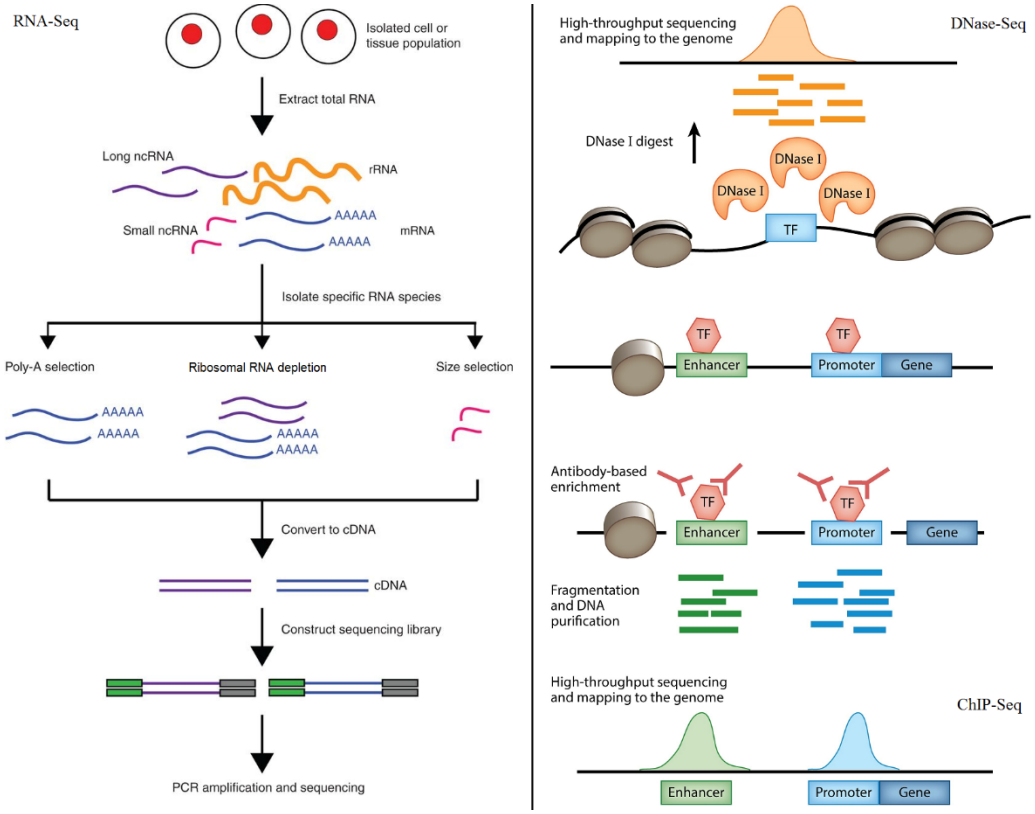


Figure 7 – Schematic overview of the RNA-Seq (left), DNase-Seq (right top), and ChIP-Seq (right, bottom) methods. In RNA-Seq, extracted RNA is enriched for the specific RNA species of interest by for example size-selection, Poly-A selection or depletion of ribosomal RNA. The RNA is converted to cDNA with attached adapters, and sequenced. DNase-Seq and ChIP-Seq is carried out by fixating the chromatin structure including DNA-bound proteins. In DNase-Seq, the chromatin is then digested by DNase I, and the protected DNA sequences are isolated and sequenced. In Chip-Seq, antibodies specific for the protein of interest are used to isolate protein-bound DNA. This DNA is then fragmented and sequenced, modified from (Chatterjee & Ahituv, 2017; Kukurba & Montgomery, 2015)

2 Results/Manuscripts

This section contains the three papers included in the thesis:

- **Paper I:** HNF4α and CDX2 regulate *YAP1* promoter activity in intestinal cell lines
- **Paper II:** CDX2 regulates interleukin-33 gene expression in intestinal epithelial cell lines
- **Paper III:** Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression

Manuscript I

2.1 Paper I HNF4α and CDX2 regulate *YAP1* promoter activity in intestinal cell lines

Sylvester Larsen[1,2], Johanne Davidsen[1,3], Katja Dahlgaard[1], Ole B. Pedersen[2], and Jesper T. Troelsen[1*]

[1]Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark; [2]Department of Clinical Immunology, Næstved Hospital, Ringstedgade 77B, 4700 Næstved, Denmark. [3]Department of Surgery, Center for Surgical Science, Zealand University Hospital, Lykkebækvej 1, 4600, Køge, Denmark

Manuscript is submitted to The Journal of Cellular Biochemistry

The aim of the study:

The aim of paper I was to investigate the role of CDX2 and HNF4α in regulating *YAP1* gene expression in colorectal cancer cell lines. The interest in *YAP1* as a CDX2 target gene was due to YAP1's role as a key modulator of the Hippo signalling pathway that is involved in colonic tissue repair and cancer development, in part, by its modulation of cell proliferation. The chromatin landscape of the *YAP1* gene was analyzed for potential CDX2 and HNF4α binding sites in the promoter and intragenic regulatory elements. The aim was to elucidate the function of the identified regulatory elements and their ability to regulate YAP1 expression through interaction with both CDX2 and HNF4α. This was done, partly, by the use of the cell model developed in paper III, where a colonic epithelial cell line lacking CDX2, but made CDX2 inducible in response to doxycycline was generated.

Key findings:

- Chromatin analysis and functional promoter assays revealed a previously unknown enhancer of the YAP1 promoter activity in the YAP1 gene, with importance for high expression levels in intestinal epithelial cells.

- Two potential CDX2 and one HNF4α binding sites were identified in the enhancer by *in-silico* transcription factor binding site analysis and protein-DNA binding was confirmed *in vitro* using electrophoretic mobility shift assay
- Western blot analysis of CDX2 inducible colon cancer epithelial cells developed in paper III, demonstrated that CDX2 increases the expression of YAP1 protein.
- CDX2 and HNF4α binding is important for the YAP1 enhancer activity in intestinal epithelial cells.

HNF4α and CDX2 regulate *YAP1* promoter activity in intestinal cell lines

Sylvester Larsen[1,2], Johanne Davidsen[1,3], Katja Dahlgaard[1], Ole B. Pedersen[2], and Jesper T. Troelsen[1*]

From the [1]Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark; [2]Department of Clinical Immunology, Næstved Hospital, Ringstedgade 77B, 4700 Næstved, Denmark. [3]Department of Surgery, Center for Surgical Science, Zealand University Hospital, Lykkebækvej 1, 4600, Køge, Denmark

Running title: *HNF4α and CDX2 regulate YAP1 promoter activity*

* To whom correspondence should be addressed: Jesper T. Troelsen: Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark;
troelsen@ruc.dk; Tel.(+45) 4674-2728

ABSTRACT

The Hippo pathway is important for tissue homeostasis, regulation of organ size and growth in most tissues. The co-transcription factor Yes-associated protein 1 (YAP1) serves as a main downstream effector of the Hippo pathway and its dysregulation increases cancer development and blocks colonic tissue repair. Nevertheless, little is known about the transcriptional regulation of *YAP1* in intestinal cells.

Bioinformatic analysis of CDX2 and HNF4α chromatin immunoprecipitated DNA from differentiated Caco-2 cells revealed a potential intragenic enhancer in the *YAP1* gene. Transfection of luciferase-expressing *YAP1* promoter reporter constructs containing the potential enhancer region validated that it is a potent enhancer of *YAP1* promoter activity in Caco-2 cells. Two potential CDX2 and one HNF4α binding sites were identified in the enhancer by *in-silico* transcription factor binding site analysis and protein-DNA binding was confirmed *in vitro* using electrophoretic mobility shift assay. It was found by chromatin immunoprecipitation experiments that CDX2 and HNF4α bind to the *YAP1* enhancer in Caco-2 cells. Finally, western blot analysis of CDX2 inducible LS174T$^{cdx2-/-}$ cells demonstrated that CDX2 increases the expression of YAP1 protein.

These results reveal a previously unknown enhancer of the *YAP1* promoter activity in the *YAP1* gene, with importance for high expression levels in intestinal epithelial cells. Additionally, CDX2 and HNF4α binding are important for the *YAP1* enhancer activity in intestinal epithelial cells.

Introduction

The Hippo pathway is a signaling pathway that is important for tissue homeostasis, regulation of organ size and tumorigenesis (Zhao et al., 2010b, Zhao et al., 2010a). It acts as a switch between proliferation and differentiation (Zhao et al., 2010b, Pan, 2010). Extra- and intracellular growth signals are relayed through the Hippo phosphorylation kinase cascade that inactivates the two downstream effectors Yes-associated protein 1 (YAP1) and Transcriptional co-activator with PDZ-binding motif (TAZ). YAP1 is a co-transcription factor that works closely with the transcription factor TEA domain family members (TEAD) to regulate the expression of many genes, especially those important for proliferation, adhesion, migration and the extracellular matrix organization (Lian et al., 2010, Stein et al., 2015). *YAP1* is expressed in all human tissues, with a relatively medium expression in the intestinal tissue (Uhlen et al., 2015).

The activity of YAP1 is affected by protein-protein interactions, phosphorylation, and re-localization of YAP1. However little is known about the regulation of *YAP1* on the transcriptional level. SOX2 has been shown to be a direct transcriptional regulator of *YAP1* in osteoprogenitor

cells, where SOX2 activates transcription of *YAP1* without being essential for *YAP1* expression (Seo et al., 2013). A study in hepatocellular carcinoma cells showed that knockdown of the heparan sulfate proteoglycan Glypican-3 decreased *YAP1* expression (Miao et al., 2013). Both ERG, ETS transcription factor (ERG) and TEA Domain Transcription Factor 4 (TEAD4) can interact with the *YAP1* promoter and increase H3K9/14Ac acetylation which leads to increased *YAP1* expression in prostate tumors (Nguyen et al., 2015). Furthermore, methylation of the *YAP1* promoter in polycystic ovary syndrome (PCOS) was also shown to increase YAP1 protein levels (Jiang et al., 2017).

As typical for key regulators of proliferation, YAP1 has been linked to cancer progression. Several studies show that dysregulation of YAP1 is important in the development of colorectal cancers (Tschaharganeh et al., 2013, Liang et al., 2014, Steinhardt et al., 2008). Furthermore, nuclear localization and overexpression of YAP1 have been correlated with a poor prognosis in colorectal cancers and other cancer types (Xia et al., 2017, Zygulska et al., 2017). Recently, YAP1 was shown to be necessary for colonic tissue repair and remodeling by mediating cell reprogramming (Yui et al., 2018).

Several signaling pathways interact within intestinal epithelial cells, adding layers of complexity to the fine-tuning of the final transcriptional regulation. The main pathways involved in the differentiation of intestinal stem cells to the various cell lineages of the gut epithelium and homeostasis maintenance is Wnt/β-catenin, NOTCH, hedgehog, Egfr and Bmp/TGFβ, reviewed in (Olsen et al., 2012, Gregorieff and Wrana, 2017). The Hippo pathway converges with several of these signaling pathways (Zhao et al., 2010a, Kriz and Korinek, 2018). Well-defined networks of intestinal epithelium-specific transcription factors control differentiation and proliferation of the intestinal epithelium (Boyd et al., 2010, Olsen et al., 2012, San Roman et al., 2015). Two key regulators in these networks are Caudal Type Homeobox 2 (CDX2) and Hepatocyte Nuclear Factor 4 alpha (HNF4α).

CDX2 is intestinal specific in adult mammals and is essential in the development and maintenance of intestinal tissue (Suh and Traber, 1996, Silberg et al., 2000). Ablation of CDX2 in *in-vitro* intestinal organoid models leads to the development of gastric-like tissue lacking intestine-specific enzymes (Simmini et al., 2014). Dysregulation of CDX2 has been linked to colorectal cancer development and affects the severity of disease (Olsen et al., 2014, Dalerba et al., 2016). Mice with decreased CDX2 expression develop malignant tumors in the colon when treated with the DNA mutagen azoxymethane (Bonhomme et al., 2003). In humans, CDX2 levels are generally lower for late-stage adenocarcinomas and tumors with low or moderately differentiation levels (Choi et al., 2006, Bakaris et al., 2008).

HNF4α is highly expressed in the intestines but is also found in the liver, stomach, pancreas, kidney, and testis (Drewes et al., 1996). HNF4α is essential for the formation of crypts in the colon and is important for the development of the intestinal epithelium (Garrison et al., 2006). Furthermore, it regulates the expression of several enzymes and regulators involved in metabolism in the adult gut epithelium (Gautier-Stein et al., 2006, Stegmann et al., 2006). HNF4α also plays a role in cancer development but it is not well understood. One study found that some HNF4α isoforms are lost in 40% of colorectal cancers, while another study found a general upregulation of HNF4α in their samples (Oshima et al., 2007, Darsigny et al., 2010).

CDX2 and HNF4α regulate each other and also auto-regulate their own expression in the intestines (Boyd et al., 2009, Xu, 1999, Saandi et al., 2013, Verzi et al., 2010a). Additionally, several other transcription factors are important in the intestinal transcription factor network regulating intestinal development, such as the GATA-binding proteins (GATA4/5/6), HNF1 homeobox proteins (HNF1A/B), and transcription factor 7 like 2 (TCF7L2) (Olsen et al., 2012). CDX2 interacts with the Wnt/β-catenin pathway and might be an active component of it (Guo et al., 2010).

Previous studies of the Hippo pathway, have predominantly focused on the protein-protein interactions of key kinases, which constitutes the core components of the pathway, and the role of the main effector proteins YAP1 and TAZ. Little is known about how the transcription of *YAP1* is controlled in intestinal tissue. This study describes a novel mechanism of transcriptional regulation of *YAP1* in intestinal cells, by the transcription factors CDX2 and HNF4α.

Results

The chromatin landscape of the YAP1 gene

To identify regulatory regions in the *YAP1* gene ChIP-seq data from differentiated Caco-2 cells were analyzed for relative abundance of CDX2 and HNF4α. The ChIP-seq data analyzed were from Caco-2 cell precipitates with antibodies for the transcription factors CDX2 and HNF4α as well as for histone acetylation H3K27Ac and methylation H3K4me2 (Verzi et al., 2010b). Data was imported as custom tracks in the UCSC Genome Browser (GRCh37/hg19 assembly), together with ENCODE datasets for DNase I activity and hepatocyte transcription factor ChIP-seq data (Sabo et al., 2006, Consortium, 2012). The *YAP1* gene was analyzed by aligning all data tracks and then identifying gene regulatory regions by comparing the signal intensity from the ChIP-seq data for transcription factors, histone modifications, and DNAseI (figure 1).

Two regions with clear CDX2 ChIP-seq peaks were chosen for further analysis, one region was in the *YAP1* promoter and another was downstream in intron 5. The chosen *YAP1* promoter region spanned 972 bp

located at hg19 chr11:101,980,493-101,981,464 (figure 1). Clear ChIP-seq peaks were detected for both CDX2, HNF4α, and H3K27Ac in the promoter region, while the H3K4me2 peak was found directly adjacent to the region. The promoter region also contains a DNAseI peak and binding sites for the intestinal transcription factors CCAAT/enhancer binding protein beta (CEBP), E1A binding protein p300 (EP300), and upstream transcription factor 1 (USF-1), indicating that this area is important for intestinal regulators.

The *YAP1* intragenic region spanned 599 bp, located at hg19 chr11:102,063,361-102,063,959 (figure 1). The intragenic region contained clear ChIP-seq signals for CDX2 and HNF4α, while both H3K4me2 and H3K27Ac peaks were found directly adjacent. The intragenic region also contains a DNAseI peak and binding sites for the intestinal transcription factors, JunD proto-oncogene (JUND), RAD21 cohesin complex component (RAD21), TCF7L2, as well as CEBP, HNF4A, EP300, and USF-1, indicating that this area is important for intestinal regulators.

CDX2 and HNF4α regulates transcription of YAP1 promoter activity

To analyze the gene-regulatory activity of the *YAP1* intragenic region we used promoter reporter assay analysis. We constructed luciferase reporter plasmids containing the *YAP1* promoter with or without the intragenic regions and transfected them into the intestinal cell lines Caco-2 and LS174T wildtype, and into HEK293 kidney cells to test whether they were able to regulate the expression of *YAP1* in intestinal cells. Cells were transfected with *YAP1* reporter plasmids with "pGL4.10-*YAP1* promoter-enhancer" or without "pGL4.10-*YAP1* promoter" the intragenic region, using an empty pGL4.10 vector as a negative control and *Lac*Z expression plasmids for transfection efficiency control.

pGL4.10 generated a very low background signal in all cell lines (figure 2). Luciferase activity of the pGL4.10-*YAP1* Promoter construct was 15-fold above background for Caco-2 and HEK293 cells, and 30-fold above background in LS174T wildtype cells, revealing that it was active in both intestinal and kidney cells.

Transfections with the pGL4.10-*YAP1* promoter-enhancer construct showed a 13-fold increase in reporter gene activity compared to *YAP1* promoter levels in Caco-2 cells while having no apparent effect in either HEK293 or LS174T wildtype cells. This showed that the intragenic region is a powerful enhancer of *YAP1* in Caco-2 cells.

Analysis of ChIP-seq data on the *YAP1* gene indicated that CDX2 and HNF4α regulate the expression of *YAP1* by binding to it. Locating specific binding sites for CDX2 and HNF4α in the *YAP1* promoter or enhancer is necessary to be able to determine if they exert their regulation by specifically binding to the DNA. The location of the binding sites was

predicted from an *in silico* analysis using the online transcription factor database JASPAR (URL: http://jaspar.genereg.net).

The prediction was carried out on the promoter and enhancer with the TFBS matrices MA04665.1 for CDX2 and MA0114.2 for HNF4α. The algorithm predicted two CDX2 sites, and two HNF4α sites in the promoter region at while three CDX2 and two HNF4α sites were found in the enhancer (Table S1).

Three sites in the *YAP1* enhancer were selected for further analysis based on their high score in the *in silico* analysis; two CDX2 sites ("CDX2-S1" and "CDX2-S2") and one HNF4α site ("HNF4α"). CDX2-S1 had the sequence CTTTATTGCCT, CDX2-S2 had the sequence ATTTATTGCAC, while the HNF4α site had the sequence CTGAACTTTGCTAAC.

We used electrophoretic mobility shift assay (EMSA) to determine whether the two CDX2 and the one HNF4α *in silico* predicted TFBS were functional and could bind endogenous proteins in Caco-2 cells. Oligonucleotides for each of the three binding sites were designed and [γ-32P]-ATP radiolabeled. Unlabeled oligonucleotides with mutated binding sequences and an unspecific competitor were used as competitors (figure 3A).

Clear specific complexes were formed on all three EMSAs when radiolabeled oligonucleotides were combined with the Caco-2 nuclear extract. Unlabeled oligonucleotides were able to compete for the binding of the radiolabeled oligonucleotides for all three sites, while the unspecific competitor was not. This indicates that the protein-oligonucleotide complexes are specific.

Two protein-DNA complexes were formed between the oligonucleotide containing the CDX2-S1 site (figure 3B) and the Caco-2 extract, while only one complex was formed using the CDX2-S2 (figure 3C) or HNF4α (figure 3D) radiolabeled oligonucleotides. Oligonucleotides with mutated binding sites were still able to form protein-oligonucleotide complexes, although not with as high affinity as wildtype oligonucleotides, showing that a low amount of binding is possible without intact binding sequences.

When antibodies to CDX2, HNF4α or a negative control Hemagglutinin (HA), were added only CDX2 and HNF4α produced supershifts, demonstrating that they are part of the protein-radiolabeled oligonucleotide complexes. In conclusion, the EMSA revealed two novel CDX2 and one HNF4α binding sites in the *YAP1* enhancer.

We wanted to investigate the effect of CDX2 and HNF4α on the *YAP1* enhancer activity in Caco-2 cells. This was done by mutating specific CDX2 and HNF4α binding sites in the pGL4.10-*YAP1*-enhancer constructs by site-directed mutagenesis (figure 4). Four constructs were created by mutation of three sites. CDX2-S1 (M1), CDX2-S2 (M2) and HNF4α (M3), and a double mutant with both M2 and M3 were

constructed and assayed for luciferase activity. The CDX2-M1 construct decreased the enhancer effect more than 5-fold, P<0.0001, while CDX2-M2 only decreased activity by 1.4-fold, P<0.01. Removing the HNF4α site (M3) caused a 2-fold decrease in activity, and the double mutant (M2+M3) lead to a 2.4-fold decreased activity from the WT P<0.001. This shows that the CDX2-S1 site is responsible for most of the enhancer activity in Caco-2 cells, while both CDX2-S2 and the HNF4α site also contributes to the effect.

Next, we performed an overexpression study, where the *YAP1* promoter and enhancer constructs were co-transfected with CDX2 and HNF4a expression plasmids to elucidate their effect on the transcriptional regulation of *YAP1* in Caco-2 cells. The relative luciferase activity was normalized to respective pGL4.10 backgrounds and the *YAP1* promoter construct. The activity of the *YAP1* promoter construct was not affected by overexpression of CDX2 (data not shown) but was increased 2-fold by HNF4α overexpression, (figure 5).

Overexpression of HNF4α with the *YAP1* enhancer construct increased the activity of the already powerful enhancer more than 6-fold, yielding more than a 75-fold difference from the promoter construct alone. It was clear that HNF4α acted as a strong activator of *YAP1* expression via the enhancer. Overexpression of CDX2 did not increase the activity of the *YAP1* enhancer construct (data not shown).

These results revealed that *YAP1* transcription was increased by HNF4α; partly through the promoter but mainly through the enhancer, while CDX2 overexpression did not increase *YAP1* promoter activity in Caco-2 cells.

CDX2 and HNF4α binding sites in the YAP1 intragenic enhancer

We performed quantitative PCR on ChIP DNA from Caco-2 cells to determine the relative abundance of CDX2 and HNF4α bound to the *YAP1* enhancer (figure 6). The relative abundance of immunoprecipitated DNA was compared to negative control DNA precipitated with hemagglutinin (HA) antibody. Primers were designed to span a small region in the *YAP1* enhancer with a clear ChIP-seq peak for CDX2 and HNF4α (Table S2B). The relative abundance of *YAP1* enhancer CDX2 immunoprecipitated DNA was 0.211 % of total input DNA, which was a more than 100-fold increase over the HA control ChIP level (0.002%). For HNF4α the relative abundance was 0.035% of total input DNA, which was an about 17-fold increase over HA control ChIP level. Both CDX2 and HNF4α binding were substantially increased compared to the negative control, which indicated, that both TF's bound to the *YAP1* enhancer in Caco-2 cells.

CDX2 induced YAP1 protein expression

Western blotting was utilized to assess whether the transcriptional regulation of *YAP1* by CDX2 seen in

the promoter assays could be translated to the protein level. Caco-2 cells are dependent on CDX2 expression and thus they are not suitable for knock-down experiments (Salari et al., 2012, Verzi et al., 2010a). However, an intestinal CDX2 knock-out cell line was recently constructed using LS174T wildtype cells (Pinto et al., 2017). The LS174T derived cells are CDX2 inducible by addition of doxycycline. Two separate LS174T derived CDX2 knock-out cells lines were constructed; CDX2 inducible LS174T$^{CDX2-/-}$ 1X, with mono-allelic knock-in of a Tet3G doxycycline-inducible CDX2 element, and CDX2 inducible LS174T$^{CDX2-/-}$ 2X with bi-allelic knock-in. We used these cell lines to clarify the regulatory role of CDX2 on the YAP1 protein expression in LS174T intestinal cells.

Caco-2, HeLa, LS174T, and the LS174T$^{CDX2-/-}$ derived cell lines were grown to confluency, proteins were extracted, and western blotting was performed. CDX2 expression was detected in both the LS174T wildtype and Caco-2 cells but not in HeLa cells (figure 7B). The YAP1 expression was high in Caco-2 cells, low in LS174T wildtype cells, and undetected in HeLa cells.

Un-induced CDX2-inducible LS174T$^{CDX2-/-}$ cell lines were negative for CDX2 expression, as expected (figure 7A). CDX2 expression in LS174T$^{CDX2-/-}$ 1X and 2X cells was induced to a much higher level than in LS174 wildtype or Caco-2 cells when stimulated with 40ng/mL doxycycline. The expression was considerably higher in 2X cells than in 1X cells. YAP1 expression in both LS174T$^{CDX2-/-}$ 1X and 2X cells was present even without CDX2 induction. This demonstrated that CDX2 expression was not essential for YAP1 expression in LS174T cells.

The YAP1 expression was increased by CDX2 induction in the LS174T$^{CDX2-/-}$ 1X and 2X cells. In LS174T$^{CDX2-/-}$ 1X cells, the YAP1 level increased to about two-fold of un-induced levels, and in the 2X cells even more. This demonstrated a dose-dependent regulation of YAP1 by CDX2 in LS174T intestinal cells.

Discussion

The Hippo pathway controls proliferation and differentiation in many tissues partly through *YAP1*. Dysregulation of *YAP1* often leads to development or exacerbation of diseases specific for the tissue in question. While the effects of *YAP1* dysregulation in the intestine has been studied, almost no papers are published on the transcriptional regulation of *YAP1*. Understanding the regulation of *YAP1* in the intestines will aid in the understanding of its role and will increase the understanding of the regulation of the Hippo pathway.

The epigenetic landscape of the *YAP1* gene and the *YAP1* promoter activity in intestinal cell lines was investigated. Through a combination of bioinformatic analysis of ChIP-seq data and transfection with luciferase promoter-reporter expression constructs with the *YAP1* promoter, we identified a powerful enhancer of

YAP1 that was functional in the intestinal cell line Caco-2.

Previous studies have reported SOX2, ERG, TEAD4, and GPC3 as regulators of *YAP1* expression in osteoprogenitor, hepatic, or prostate tissue. While some detected epigenetic changes to the *YAP1* promoter, none of them identified intragenic regulatory regions or specific binding sites in the *YAP1* gene (Seo et al., 2013, Miao et al., 2013, Nguyen et al., 2015, Jiang et al., 2017). Here we presented the first clear evidence of an enhancer in the *YAP1* gene that is active in intestinal cells.

We measured the importance of HNF4α and CDX2 for the *YAP1* promoter activity in intestinal cell lines. By overexpression experiments in Caco-2 cells, it was determined that HNF4α was an activator of *YAP1* expression (figure 5). Promoter reporter assays with mutated CDX2 and HNF4α binding sites in the *YAP1* enhancer showed that these sites were important for the enhancer activity (figure 4).

HNF4α and CDX2 are essential components of the intestinal transcription factor network that controls development and homeostasis, and CDX2 is normally only expressed in the intestines after birth in mammals (James and Kazenwadel, 1991, Suh et al., 1994). During intestinal differentiation, CDX2 and HNF4α are a part of a transcription factor network that tightly regulates a large array of intestinally expressed genes (Olsen et al., 2012, Boyd et al., 2009, Boyd et al., 2010). *YAP1* is expressed in several tissues but is especially highly expressed in colon cancers (Steinhardt et al., 2008, Liang et al., 2014). We suggest that CDX2 and HNF4α adds an intestinal-specific layer of control to the Hippo pathway through transcriptional regulation of *YAP1*.

CDX2 protein expression was induced in intestinal cell lines to determine how the CDX2 protein level affects YAP1 protein expression. Using western blotting to measure YAP1 and CDX2 protein expression in CDX2 induced LS174T$^{CDX2-/-}$ cell lines, we found that CDX2 is not necessary for YAP1 expression, however, induction of CDX2 increased YAP1 expression. It was somewhat unexpected that overexpression of CDX2 so clearly increased YAP1 protein expression cells since our results show that the *YAP1* enhancer is inactive in LS174T cells (figure 2). However, the transfected promoter-enhancer constructs only contained parts of the full *YAP1* gene, and thus other regulatory elements than the cloned promoter and enhancer could be responsible for the upregulation of YAP1 protein expression by CDX2. Indeed, two additional regions in *YAP1* also contained enriched CDX2 and HNF4a ChIP-Seq signals, one in intron 1 and one in intron 4 (figure 1). It is possible that CDX2 acts through one or both of these regions to increase YAP1 protein levels seen on the western blot (figure 7).

We were not able to induce YAP1 expression back to the endogenous LS174T wildtype level using our CDX2 inducible

LS174T$^{CDX2-/-}$ derived cells lines. Previous work has established this pattern for almost all CDX2 target genes that are re-induced after CDX2 knockout (Pinto et al., 2017). We hypothesize that the disruption of major transcriptional pathways by knockout of CDX2 causes non-reversible changes to the cell.

By constructing HNF4α knockout intestinal cells lines, it would be possible to determine how important HNF4α is for the general transcription of *YAP1* and its precise contribution to the enhancer effect. Although this might not be possible since it has not been possible to find any published work describing an HNF4α knockout intestinal cell line and because our lab has not been able to create one either.

In summary, we have identified gene regulatory regions in the *YAP1* promoter and an enhancer with importance for high expression levels in Caco-2 cells. We found that that CDX2 and HNF4α are important for *YAP1* enhancer activity.

Experimental procedures

Cultivating human cell lines

Caco-2 and LS174T wildtype, colon adenocarcinoma cell line; HEK293, human embryonic kidney cell line; HeLa, cervical adenocarcinoma cell line; LS174T derived CDX2 knockout (LS174T$^{CDX2-/-}$) with Tet-on inducible CDX2 expression cell line (Pinto et al., 2017) were grown in T175 culture flasks in Dulbecco's Modified Eagle Medium (Lonza), added 10% Fetal Bovine Serum gold (PAA) and 100 U/mL Penicillin-Streptomycin. Cells were incubated at 37 °C in 5 % CO_2 and passaged every 3-4 days when ~80 % confluent. Passaging was done by removing media, rinsing three times with 85 mM sodium citrate, adding 1mL 0.05 % trypsin EDTA (Invitrogen) and incubating 3-5 minutes at 37 °C in 5 % CO_2. Cells were seeded at 0.75*10^6 (Caco-2), 1.00*10^6 (HEK293 and HeLa), and 3.00*10^6 (all LS174T based cell lines) in a T75 flask with 15 mL Media.

Western Blotting

Protein extraction was carried out on cells seeded in 6-well plates at 100,000/well for Caco-2, 300,000/well for LS174T wildtype and LS174T$^{CDX2-/-}$ cells. After 48hrs, the media was changed to either Doxycycline (DOX) free media or media with varying concentration of DOX. At 72hrs cells were rinsed with cold PBS and lysed for 15 minutes with 100µl/well 1X RIPA lysis buffer (1X PBS, 300mM NaCl, 1% Tergitol NP-40, 0.1% SDS, 0.5% 7-Deoxycholic acid sodium salt, 0.5µM EDTA pH 8.0) with freshly added 1mM DTT and 2µL/mL protease inhibitor mix p8340 (Sigma-Aldrich). Lysates were centrifuged for 20 minutes at 12.000g at 4°C and supernatant saved at -20°C. Protein concentration was determined by Bradford analysis (Bio-Rad).

10µg or 20µg protein was mixed 1:4 (v/v) with Bolt loading buffer and 1:10 (v/v) with Bolt sample reducing agent. Samples were incubated at 70°C for 10min and

loaded on a Bolt 4-12% Bis-Tris Plus gel (Thermo Fisher Scientific) along with prestained protein marker PageRuler (Thermo Fisher Scientific). SDS-PAGE was performed in 1X Bolt MOPS running buffer, (Thermo Fisher Scientific) for 50min at 150V. Gels were transferred by wet-electrotransfer to PVDF membranes for 75min at 25V and 100mA in 1X NuPage transfer buffer, (Thermo Fisher Scientific). Membranes were blocked with dry skim milk diluted to 5% in Wash buffer (1X PBS with 0.1% Tween-20) for 1 hour at room temperature, washed with Wash buffer 3 x 7 min and incubated ON at 4°C with primary antibody diluted in Dilution buffer (2.5% skim milk in Wash Buffer). Membranes were washed 3 x 7 min and incubated with secondary antibody for 1 hour at RT and washed 3 x 7 min. Bands were visualized by incubating with the ECL solution SuperSignal™ West Dura Extended Duration Substrate (Thermo Fisher Scientific). Antibodies used: CDX2 1:2000, (Biogenex, MU392A-UC; YAP1 1:15000, (Abcam, ab52771); GAPDH 1:30000 (Fitzgerald, 10R-G109a), Goat anti-rabbit HRP 1:4000 (Life Technologies, 32260); Goat anti-mouse HRP 1:10000 (Life Technologies, 32230).

Construction of luciferase reporter constructs

The *YAP1* promoter and enhancer sequence were PCR amplified and gel purified with primers ordered from Eurofins Genomics (Table S2A). pGL4.10 vector (Promega) was digested with HindIII and gel purified. In-Fusion cloning (Clontech) was carried out according to manufacturer's protocol, using gel-purified linearized pGL4.10-vectors and promoter insert in a 1:2 molar ratio. 2.5µl cloning reaction was used for transformation of One Shot TOP10 chemically competent E. *coli* (Thermo Fisher Scientific). Colonies were isolated and sequenced (Beckman Coulter Genomics). For the construction of the pGL4.10-Yap1-promoter+enhancer plasmid, the pGL4.10-Yap1-promoter plasmid was digested with SalI and gel purified, and In-Fusion cloned with enhancer insert using the same procedure.

Construction of reporter constructs with mutated binding sites

pGL4.10-Yap1-promoter+enhancer fragments with mutated binding sites were created by PCR amplifying forward primers containing mutated binding sites, with reverse primers containing In-fusion tails and vice versa. The products were subsequently combined in a third PCR amplification. Mutated inserts were In-Fusion cloned and transformed. Plasmids were purified and digested with *Pst*I, *Xho*I, or *Xba*I and the fragments were gel purified and sequenced.

Transfection of cell lines

Cells were seeded in 24 well plates at 50,000 cells/well (Caco-2) or at 100,000 cells/well (LS714T and CDX2 inducible LS174T$^{CDX2-/-}$) and transfected after 24 hrs. For each well, a transfection mixture of 50 µl

DNA/PEI mix was prepared, consisting of 2 µM Polyethyleneimine (PEI) (Alfa Aesar) diluted in 150 mM NaCl to 25 µl and 300 ng plasmid DNA diluted with 150 mM NaCl to 25 µl. The transfection mix contained 25 ng CMV promoter-driven overexpression plasmid or pCDNA3.1+ containing a CMV promoter, to equalize molar ratio of CMV promoter constructs, 50 ng promoter reporter plasmid, 25 ng CMV-LacZ plasmid for internal control of transfection efficiency, and 200 ng pBluescript SK+ II used as inactive DNA to reach 300 ng DNA/well. Volume was adjusted with 150 mM NaCl and transfection mix was incubated 60 min at RT. The transfection mix was added in a dropwise fashion to wells, followed by light shaking. After transfection, plates were centrifuged at 390 g for 5 min and incubated overnight at 37 °C and 5 % CO_2. The media was changed after 24 hrs, and cells were grown for an additional 24 hrs before assayed for luciferase and β-galactosidase activity.

Measuring luciferase activity of reporter constructs

Measurements were carried out on a GloMax® 96 Microplate Luminometer using the Dual-Light™ Luciferase & β-Galactosidase Reporter Gene Assay System (Thermo Fisher Scientific). Cells were rinsed three times with 1X PBS and lysed with 130 µL TROPIX lysis solution containing 0.5 mM DTT and incubated for 10 min on ice. 10 µL lysate was transferred to a 96 well GloMax Luminometer Light Plate, and luciferase activity was measured using a 5-sec integration time and a 2-sec delay. Beta-galactosidase activity was measured 45 minutes after luciferase activity.

Chromatin immunoprecipitation assay and Quantitative PCR

Two separate immunoprecipitations from Caco-2 cells were generated. Both were prepared following the protocol described in (Coskun et al., 2012). The first set included CDX2 and Hemagglutinin (HA) precipitations, where the second further contained an HNF4α precipitation, using the antibodies; CDX2 (BioGenex, MU392A-UC), HA (Santa Cruz Biotechnology, Clone Y-11 X, SC-805 X), or HNF4a (Santa Cruz Biotechnology, Clone SC 171 X, SC-8987 X). qPCR was used to detect binding of CDX2, HNF4α, and HA to the *YAP1* promoter and enhancer in the samples, using primers from Eurofins Genomics (Table S2B). qPCR was carried out with SsoFast EvaGreen Supermix with Low ROX (Bio-Rad) on 1 µl ($1/20^{th}$ of total IP sample DNA) ChIP DNA or H_2O with 0.5 µM primers on a Stratagene MX3005P real-time thermal cycler (Agilent Technologies). Relative quantification of ChIP-DNA was calculated as a percentage of the input DNA by the delta-delta method (Livak and Schmittgen, 2001).

Electrophoretic mobility shift assay

Designing and annealing of oligonucleotides

EMSA probes used was oligonucleotides with 5' overhangs, containing either the wildtype binding sequences, an unspecific competitor sequence, or specific competitor sequences with mutated binding sites (figure 3A). 250 pmol of each complementary oligonucleotides were annealed in a total volume of 100µL 0.1M NaCl by heating to 95 °C followed by passive cooling until room temperature.

For the preparation of radiolabeled oligonucleotides 2.5 pmol annealed oligonucleotide pairs were mixed with 0.5 µL T4 Polynucleotide Kinase, 1 µL 10X forward kinase buffer, 5 µL [γ-32P]-ATP (3000Ci/mmol 5 mCi/ml EasyTide Lead (#NEG502H250UC, PerkinElmer) and 2.5 µL H_2O for a total volume of 10 µL. The reaction was incubated at 37 °C for 30 minutes after which 20 µL 1X TE buffer was added. The probes were purified with Illustra MicroSpin G-25 Columns (GE Healthcare Life Sciences) and diluted with 1X TE buffer to a total volume of 100 µL and a concentration of about 25 fmol/µL. Radiolabeled oligonucleotides were further diluted 10X in TE buffer immediately before loading.

Caco-2 nuclear extracts used for the EMSA was prepared as previously described (Troelsen et al., 2003). The EMSA reaction contained 1 µl differentiated Caco-2 nuclear extract 4 µl dialysis buffer (25 mM Hepes pH 7.6, 0.1 mM EDTA, 40 nM KCl, and 10 % glycerol), 10 µl Gel-shift buffer (25mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 5 % Ficoll 400, 2.5% glycerol, 60 mM KCl, 0.5 mM EDTA, 1 mM DTT, and protease inhibitors), and 0.5 µl dI-dC (homopolymer of deoxyinosine and deoxycytidine). For competition 1 µL 250 fmol/µL unlabeled oligonucleotides, either unspecific, wild-type, or mutated oligonucleotides was added. Supershift assays contained 1 µL antibody; CDX2 (BioGenex, MU392A-UC), HA (Santa Cruz Biotechnology, Clone Y-11 X, SC-805 X), or HNF4a (Santa Cruz Biotechnology, Clone SC 171 X, SC-8987 X). The reaction was incubated 20 min on ice, and then added 1 µL 2.5 fmol of ^{32}-P labeled probe followed by 20 min incubation on ice.

Before gel-loading, 2 µl Gel-shift loading buffer (10 % glycerol, 0.2 % Bromphenol blue and 0.5X Tris-borate-EDTA buffer (2X TBE, 44.5 mM Tris-HCl pH 8.0, 1 mM EDTA, and 44.5 mM boric acid)) was added to the reaction mix. The reaction mix was loaded on a precooled non-denaturing 5% polyacrylamide gel (1 gel: 2.25 mL 30% Acrylamide/Bis-acrylamide (29:1), 0.67mL 10X TBE, 0.78 mL 87% glycerol, 9.71 mL H_2O, 54µL 25% ammonium persulfate (AMPS), and 6.9µL TEMED), using precooled 0.5X TBE as running buffer. Runtime was 45 minutes at 100 mV and 25 mA/gel with active cooling. The gel was dried on a slab gel dryer for two hours and exposed on a phosphor-imager for 24 hours. The phosphor screen was scanned on a

Storm 840 scanner (Molecular Dynamics) and the image was processed using the Image-Quant Software version 5.2.

Statistics

Where applicable, values are represented by mean values with 95% confidence intervals or SD values. The P-values have been determined by two-tailed Student's t-tests or one-way ANOVA using GraphPad PRISM v7.0. The significance levels are shown as * P< 0.05, ** P< 0.01, *** P< 0.001, **** P< 0.0001

Acknowledgements: We would like to thank Mehmet Coskun for providing one set of immunoprecipitated DNA from Caco-2 cells.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

References

BAKARIS, S., CETINKAYA, A., EZBERCI, F. & EKERBICER, H. 2008. Expression of homeodomain protein CDX2 in colorectal adenoma and adenocarcinoma. *Histology and histopathology,* 23**,** 1043-7.

BONHOMME, C., DULUC, I., MARTIN, E., CHAWENGSAKSOPHAK, K., CHENARD, M. P., KEDINGER, M., BECK, F., FREUND, J. N. & DOMON-DELL, C. 2003. The Cdx2 homeobox gene has a tumour suppressor function in the distal colon in addition to a homeotic role during gut development. *Gut,* 52**,** 1465-71.

BOYD, M., BRESSENDORFF, S., MØLLER, J., OLSEN, J. & TROELSEN, J. T. 2009. Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC gastroenterology,* 9**,** 68-68.

BOYD, M., HANSEN, M., JENSEN, T. G. K., PEREARNAU, A., OLSEN, A. K., BRAM, L. L., BAK, M., TOMMERUP, N., OLSEN, J. & TROELSEN, J. T. 2010. Genome-wide analysis of CDX2 binding in intestinal epithelial cells (Caco-2). *The Journal of biological chemistry,* 285**,** 25115-25.

CHOI, B. J., KIM, C. J., CHO, Y. G., SONG, J. H., KIM, S. Y., NAM, S. W., LEE, S. H., YOO, N. J., LEE, J. Y. & PARK, W. S. 2006. Altered expression of CDX2 in colorectal cancers. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica,* 114**,** 50-4.

CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

COSKUN, M., OLSEN, A. K., HOLM, T. L., KVIST, P. H., NIELSEN, O. H., RIIS, L. B., OLSEN, J. & TROELSEN, J. T. 2012. TNF-α-induced down-regulation of CDX2 suppresses MEP1A expression in colitis. *Biochimica et biophysica acta,* 1822**,** 843-51.

DALERBA, P., SAHOO, D., PAIK, S., GUO, X., YOTHERS, G., SONG, N., WILCOX-FOGEL, N., FORGO, E., RAJENDRAN, P. S., MIRANDA, S. P., HISAMORI, S., HUTCHISON, J., KALISKY, T., QIAN, D., WOLMARK, N., FISHER, G. A., VAN DE RIJN, M. & CLARKE, M. F. 2016. CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med,* 374**,** 211-22.

DARSIGNY, M., BABEU, J. P., SEIDMAN, E. G., GENDRON, F. P., LEVY, E., CARRIER, J., PERREAULT, N. & BOUDREAU, F. 2010. Hepatocyte nuclear factor-4alpha promotes gut neoplasia in mice and protects against the production of reactive oxygen species. *Cancer Res,* 70**,** 9423-33.

DREWES, T., SENKEL, S., HOLEWA, B. & RYFFEL, G. U. 1996. Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Molecular and cellular biology,* 16**,** 925-31.

GARRISON, W. D., BATTLE, M. A., YANG, C., KAESTNER, K. H., SLADEK, F. M. & DUNCAN, S. A. 2006. Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. *Gastroenterology,* 130**,** 1207-20.

GAUTIER-STEIN, A., ZITOUN, C., LALLI, E., MITHIEUX, G. & RAJAS, F. 2006. Transcriptional regulation of the glucose-6-phosphatase gene by cAMP/vasoactive intestinal peptide in the intestine. Role of HNF4alpha, CREM, HNF1alpha, and C/EBPalpha. *The Journal of biological chemistry,* 281**,** 31268-78.

GREGORIEFF, A. & WRANA, J. L. 2017. Hippo signalling in intestinal regeneration and cancer. *Current Opinion in Cell Biology,* 48**,** 17-25.

GUO, R.-J., FUNAKOSHI, S., LEE, H. H., KONG, J. & LYNCH, J. P. 2010. The intestine-specific transcription factor Cdx2 inhibits β-catenin/TCF transcriptional activity by disrupting the β-catenin–TCF protein complex. *Carcinogenesis,* 31**,** 159-166.

JAMES, R. & KAZENWADEL, J. 1991. Homeobox gene expression in the intestinal epithelium of adult mice. *The Journal of biological chemistry,* 266**,** 3246-51.

JIANG, L.-L., XIE, J.-K., CUI, J.-Q., WEI, D., YIN, B.-L., ZHANG, Y.-N., CHEN, Y.-H., HAN, X., WANG, Q. & ZHANG, C.-L. 2017. Promoter methylation of yes-associated protein (YAP1) gene in polycystic ovary syndrome. *Medicine,* 96**,** e5768-e5768.

KRIZ, V. & KORINEK, V. 2018. Wnt, RSPO and Hippo Signalling in the Intestine and Intestinal Stem Cells. *Genes (Basel),* 9.

LIAN, I., KIM, J., OKAZAWA, H., ZHAO, J., ZHAO, B., YU, J., CHINNAIYAN, A., ISRAEL, M. A., GOLDSTEIN, L. S., ABUJAROUR, R., DING, S. & GUAN, K. L. 2010. The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev,* 24**,** 1106-18.

LIANG, K., ZHOU, G., ZHANG, Q., LI, J. & ZHANG, C. 2014. Expression of hippo pathway in colorectal cancer. *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association,* 20**,** 188-94.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.),* 25**,** 402-8.

MATHELIER, A., FORNES, O., ARENILLAS, D. J., CHEN, C. Y., DENAY, G., LEE, J., SHI, W., SHYR, C., TAN, G., WORSLEY-HUNT, R., ZHANG, A. W., PARCY, F., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2016. JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research,* 44**,** D110-D115.

MIAO, H.-L., PAN, Z.-J., LEI, C.-J., WEN, J.-Y., LI, M.-Y., LIU, Z.-K., QIU, Z.-D., LIN, M.-Z., CHEN, N.-P. & CHEN, M. 2013. Knockdown of GPC3 inhibits the proliferation of Huh7 hepatocellular carcinoma cells through down-regulation of YAP. *Journal of cellular biochemistry,* 114**,** 625-31.

NGUYEN, L. T., TRETIAKOVA, M. S., SILVIS, M. R., LUCAS, J., KLEZOVITCH, O., COLEMAN, I., BOLOURI, H., KUTYAVIN, V. I., MORRISSEY, C., TRUE, L. D., NELSON, P. S. & VASIOUKHIN, V. 2015. ERG Activates the

YAP1 Transcriptional Program and Induces the Development of Age-Related Prostate Tumors. *Cancer Cell,* 27**,** 797-808.

OLSEN, A. K., BOYD, M., DANIELSEN, E. T. & TROELSEN, J. T. 2012. Current and emerging approaches to define intestinal epithelium-specific transcriptional networks. *American journal of physiology. Gastrointestinal and liver physiology,* 302**,** G277-86.

OLSEN, J., ESPERSEN, M. L., JESS, P., KIRKEBY, L. T. & TROELSEN, J. T. 2014. The clinical perspectives of CDX2 expression in colorectal cancer: a qualitative systematic review. *Surg Oncol,* 23**,** 167-76.

OSHIMA, T., KAWASAKI, T., OHASHI, R., HASEGAWA, G., JIANG, S., UMEZU, H., AOYAGI, Y., IWANARI, H., TANAKA, T., HAMAKUBO, T., KODAMA, T. & NAITO, M. 2007. Downregulated P1 promoter-driven hepatocyte nuclear factor-4alpha expression in human colorectal carcinoma is a new prognostic factor against liver metastasis. *Pathology international,* 57**,** 82-90.

PAN, D. 2010. The hippo signaling pathway in development and cancer. *Developmental cell,* 19**,** 491-505.

PINTO, R., HANSEN, L., HINTZE, J., ALMEIDA, R., LARSEN, S., COSKUN, M., DAVIDSEN, J., MITCHELMORE, C., DAVID, L., TROELSEN, J. T. & BENNETT, E. P. 2017. Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression. *Nucleic acids research,* 40**,** 439-452.

SABO, P. J., KUEHN, M. S., THURMAN, R., JOHNSON, B. E., JOHNSON, E. M., CAO, H., YU, M., ROSENZWEIG, E., GOLDY, J., HAYDOCK, A., WEAVER, M., SHAFER, A., LEE, K., NERI, F., HUMBERT, R., SINGER, M. A., RICHMOND, T. A., DORSCHNER, M. O., MCARTHUR, M., HAWRYLYCZ, M., GREEN, R. D., NAVAS, P. A., NOBLE, W. S. & STAMATOYANNOPOULOS, J. A. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods,* 3**,** 511-8.

SALARI, K., SPULAK, M. E., CUFF, J., FORSTER, A. D., GIACOMINI, C. P., HUANG, S., KO, M. E., LIN, A. Y., VAN DE RIJN, M. & POLLACK, J. R. 2012. CDX2 is an amplified lineage-survival oncogene in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** E3196-E3205.

SAN ROMAN, A. K., ARONSON, B. E., KRASINSKI, S. D., SHIVDASANI, R. A. & VERZI, M. P. 2015. Transcription Factors GATA4 and HNF4A Control Distinct Aspects of Intestinal Homeostasis in Conjunction with Transcription Factor CDX2. *Journal of Biological Chemistry,* 290**,** 1850-1860.

SEO, E., BASU-ROY, U., GUNARATNE, P. H., COARFA, C., LIM, D.-S. S., BASILICO, C. & MANSUKHANI, A. 2013. SOX2 regulates YAP1 to maintain stemness and determine cell fate in the osteo-adipo lineage. *Cell Reports,* 3**,** 2075-2087.

SILBERG, D. G., SWAIN, G. P., SUH, E. R. & TRABER, P. G. 2000. Cdx1 and cdx2 expression during intestinal development. *Gastroenterology,* 119**,** 961-71.

SIMMINI, S., BIALECKA, M., HUCH, M., KESTER, L., VAN DE WETERING, M., SATO, T., BECK, F., VAN OUDENAARDEN, A., CLEVERS, H. & DESCHAMPS, J. 2014. Transformation of intestinal stem cells into gastric stem cells on loss of transcription factor Cdx2. *Nature Communications,* 5**,** 5728-5728.

STEGMANN, A., HANSEN, M., WANG, Y., LARSEN, J. B., LUND, L. R., RITIE, L., NICHOLSON, J. K., QUISTORFF, B., SIMON-ASSMANN, P., TROELSEN, J. T. & OLSEN, J. 2006. Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation. *Physiol Genomics,* 27**,** 141-55.

STEIN, C., BARDET, A. F., ROMA, G., BERGLING, S., CLAY, I., RUCHTI, A., AGARINIS, C., SCHMELZLE, T., BOUWMEESTER, T., SCHUBELER, D. & BAUER, A. 2015. YAP1 Exerts Its Transcriptional Control via TEAD-Mediated Activation of Enhancers. *PLoS Genet,* 11**,** e1005465.

STEINHARDT, A. A., GAYYED, M. F., KLEIN, A. P., DONG, J., MAITRA, A., PAN, D., MONTGOMERY, E. A. & ANDERS, R. A. 2008. Expression of Yes-associated protein in common solid tumors. *Human pathology,* 39**,** 1582-9.

SUH, E., CHEN, L., TAYLOR, J. & TRABER, P. G. 1994. A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol Cell Biol,* 14**,** 7340-51.

SUH, E. & TRABER, P. G. 1996. An intestine-specific homeobox gene regulates proliferation and differentiation. *Molecular and cellular biology,* 16**,** 619-25.

SAANDI, T., BARAILLE, F., DERBAL-WOLFROM, L., CATTIN, A.-L., BENAHMED, F., MARTIN, E., CARDOT, P., DUCLOS, B., RIBEIRO, A., FREUND, J. N. & DULUC, I. 2013. Regulation of the tumor suppressor homeogene Cdx2 by HNF4α in intestinal cancer. *Oncogene,* 32**,** 3782-8.

TROELSEN, J. T., MITCHELMORE, C. & OLSEN, J. 2003. An enhancer activates the pig lactase phlorizin hydrolase promoter in intestinal cells. *Gene,* 305**,** 101-111.

TSCHAHARGANEH, D. F., CHEN, X., LATZKO, P., MALZ, M., GAIDA, M. M., FELIX, K., LADU, S., SINGER, S., PINNA, F., GRETZ, N., STICHT, C., TOMASI, M. L., DELOGU, S., EVERT, M., FAN, B., RIBBACK, S., JIANG, L., BROZZETTI, S., BERGMANN, F., DOMBROWSKI, F., SCHIRMACHER, P., CALVISI, D. F. & BREUHAHN, K. 2013. Yes-associated protein up-regulates Jagged-1 and activates the Notch pathway in human hepatocellular carcinoma. *Gastroenterology,* 144**,** 1530-1542.e12.

UHLEN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJOSTEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P. H., BERLING, H., TEGEL, H.,

MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN, M., VON HEIJNE, G., NIELSEN, J. & PONTEN, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.

VERZI, M. P., HATZIS, P., SULAHIAN, R., PHILIPS, J., SCHUIJERS, J., SHIN, H., FREED, E., LYNCH, J. P., DANG, D. T., BROWN, M., CLEVERS, H., LIU, X. S. & SHIVDASANI, R. A. 2010a. TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 15157-62.

VERZI, M. P., SHIN, H., HE, H. H., SULAHIAN, R., MEYER, C. A., MONTGOMERY, R. K., FLEET, J. C., BROWN, M., LIU, X. S. & SHIVDASANI, R. A. 2010b. Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Developmental cell,* 19**,** 713-26.

XIA, J., ZENG, M., ZHU, H., CHEN, X., WENG, Z. & LI, S. 2017. Emerging role of Hippo signalling pathway in bladder cancer. *J Cell Mol Med*.

XU, F. 1999. Cell Type-specific Autoregulation of the Caudal-related Homeobox Gene Cdx-2/3. *Journal of Biological Chemistry,* 274**,** 34310-34316.

YUI, S., AZZOLIN, L., MAIMETS, M., PEDERSEN, M. T., FORDHAM, R. P., HANSEN, S. L., LARSEN, H. L., GUIU, J., ALVES, M. R. P., RUNDSTEN, C. F., JOHANSEN, J. V., LI, Y., MADSEN, C. D., NAKAMURA, T., WATANABE, M., NIELSEN, O. H., SCHWEIGER, P. J., PICCOLO, S. & JENSEN, K. B. 2018. YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell,* 22**,** 35-49 e7.

ZHAO, B., LI, L. & GUAN, K. L. 2010a. Hippo signaling at a glance. *J Cell Sci,* 123**,** 4001-6.

ZHAO, B., LI, L., LEI, Q. & GUAN, K. L. 2010b. The Hippo–YAP pathway in organ size control and tumorigenesis: An updated version. *Genes & development,* 24**,** 862-874.

ZYGULSKA, A. L., KRZEMIENIECKI, K. & PIERZCHALSKI, P. 2017. Hippo pathway - brief overview of its relevance in cancer. *J Physiol Pharmacol,* 68**,** 311-335.

FOOTNOTES:

List of abbreviations used:
YAP1, Yes-associated protein 1; TAZ, Transcriptional co-activator with PDZ-binding motif; TEAD, Transcription factor TEA domain family members; ERG, ERG, ETS transcription factor; TEAD4, TEA Domain Transcription Factor 4; PCOS, Polycystic ovary syndrome; GATA4/5/6, GATA-binding proteins 4/5/6; HNF1A/B, HNF1 homeobox proteins; TCF7L2, Transcription factor 7 like 2; HNF4α, Hepatocyte Nuclear Factor 4 alpha; CDX2, Caudal Type Homeobox 2; USF-1, Upstream transcription factor 1; CEBP, CCAAT/enhancer binding protein beta; EP300, E1A binding protein p300; JUND, JunD proto-oncogene; RAD21, RAD21 cohesin complex component; EMSA, Electrophoretic mobility shift assay; HA, Hemagglutinin; DOX, Doxycycline; PEI, Polyethyleneimine; AMPS, Ammonium persulfate.

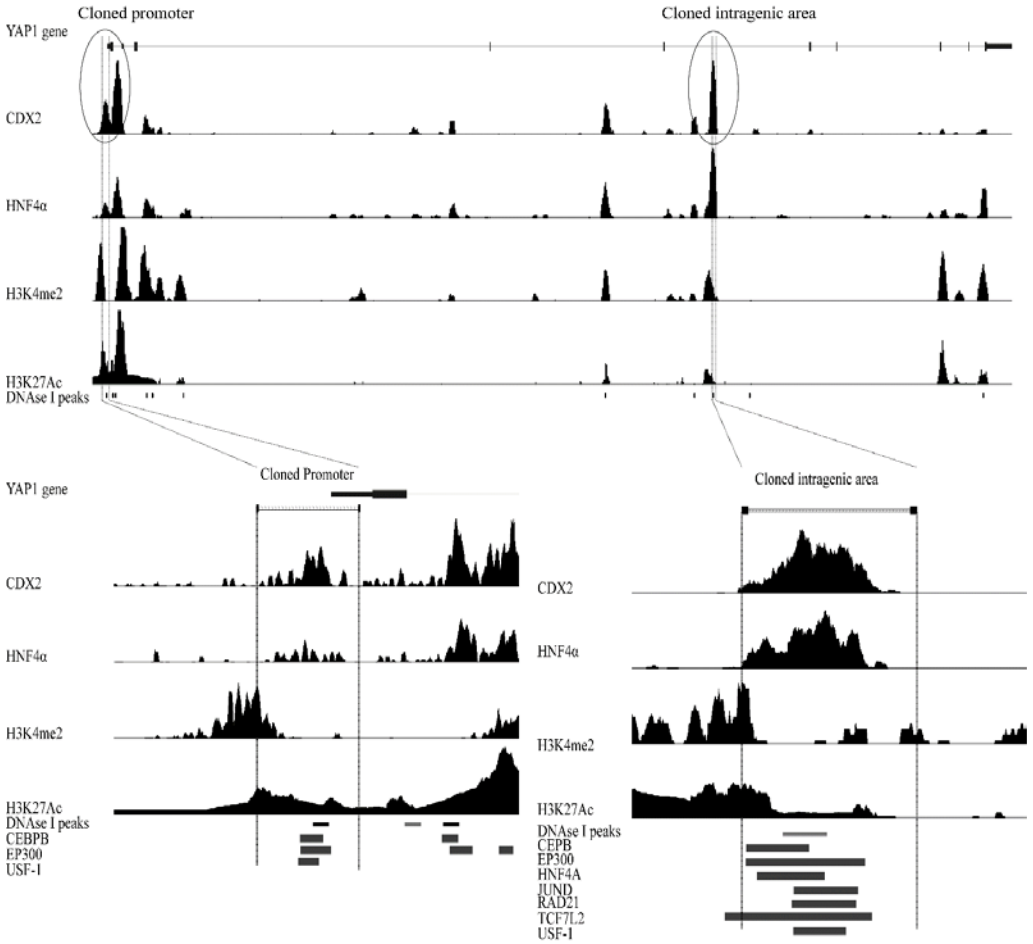


Figure 1 – Epigenetic analysis of the *YAP1* gene. **TOP)** View of the *YAP1* gene (NM_001130145) in GRCh37/hg19, region (chr11:101,979,192-102,104,154). ChIP-seq data for CDX2, HNF4α, and H3K4me2 from confluent Caco-2 cells is shown as density graphs (Verzi et al., 2010a). Circles and dashed lines indicate the promoter and an intragenic regulatory region in the *YAP1* gene. **Bottom Left)** View of the promoter region of the *YAP1* gene (NM_001130145) in GRCh37/hg19 spanning 972bp(chr11:101,979,159-101,982,952). The ENCODE Transcription Factor ChIP-seq Peaks shows selected data from the hepatocyte cell line HEPG2 (Consortium, 2012). **Bottom right)** View of the intragenic regulatory region in intron 5 of the *YAP1* gene in GRCh37/hg19 spanning 599bp (chr11:102,062,986-102,064,334). DNaseI peak track was published by (Sabo et al., 2006).

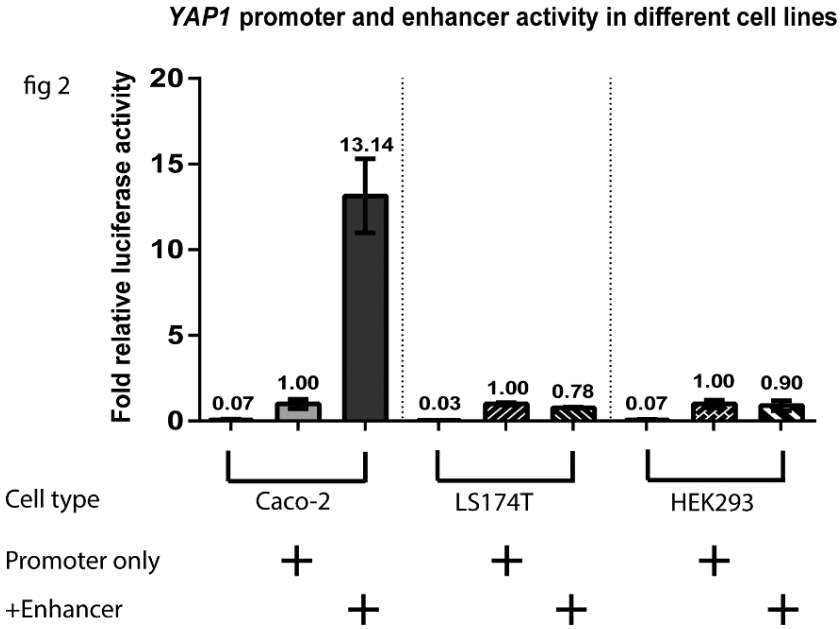


Figure 2 – pGL4-10 Luciferase reporter constructs containing *YAP1* promoter and enhancer were transfected into Caco-2, LS174T wildtype, and HEK293 cells. Their relative luciferase/beta-galactosidase activity was measured in cell lysates using a Promega dual light assay. The *YAP1* promoter is active in all three cell lines. Addition of the *YAP1* enhancer increases activity more than 13-fold in Caco-2 while having negligible or no effect in LS174T wildtype and HEK293 cells. Vertical lines separate data from independent assays. Bars are mean values, with error bars showing 95% CI. $P<0.001$, $n=4$)

A

Oligos	Sequences 5´-> 3´
Unspecific competitor	5´AACGTAGCTGATCGAATCGGTTAC 3´TGCATCGACTAGCTTAGCCAATGA
CDX2-S1	5´TTTTGAACTATTTATTGCACCATT 3´AAACTTGATAAATAACGTGGTAAA
CDX2-S1-Mut	5´AGATACTTTTGAACT**CTGCAG**GCACCATTTAAGTT 3´CTATGAAAACTTGA**GACGTC**CGTGGTAAATTCAAA
CDX2-S2	5´AAGTTCAGCTTTATTGCCTTAAAG 3´TCAAGTCGAAATAACGGAATTTCT
CDX2-S2-Mut	5´TTAGCAAAGTTCAGC**CTCGAG**GCCTTAAAGATGAT 3´ATCGTTTCAAGTCG**GAGCTC**CGGAATTTCTACTAC
HNF4α	5´TTCATGTTAGCAAAGTTCAGCTTT 3´AGTACAATCGTTTCAAGTCGAAAT
HNF4α-Mut	5´ACATGTTCATGTTAG**TCTAGA**TCAGCTTTATTGCC 3´GTACAAGTACAATC**AGATCT**AGTCGAAATAACGGA

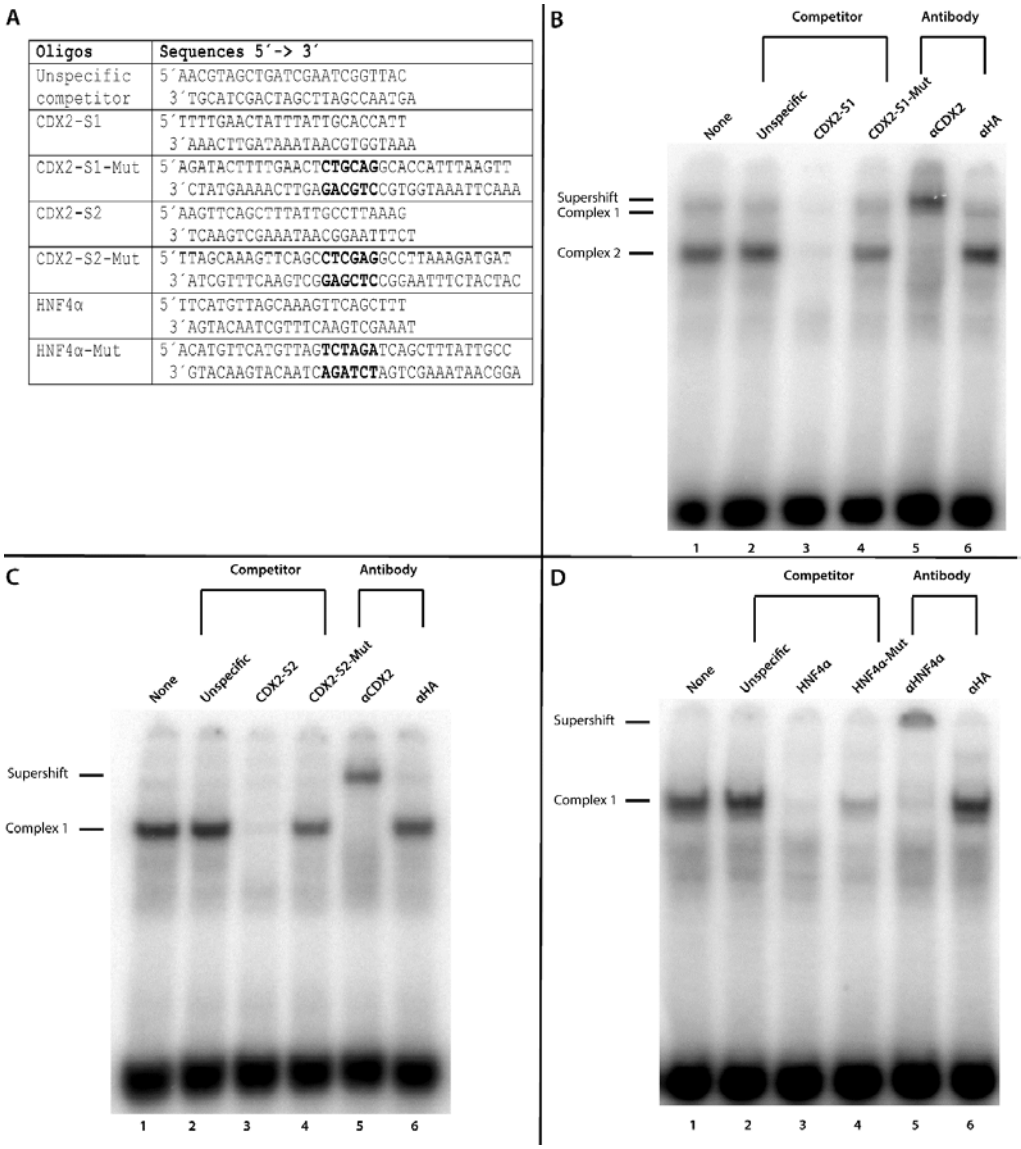


Figure 3 – Electrophoretic mobility shift assay of two CDX2 sites and one HNF4α site in the *YAP1* enhancer, annotated CDX2-S1 and CDX2-S2 and HNF4α. **A)** Table of double-stranded oligonucleotides with mutated binding sites in bold. **B)** All wells contain: 1 µL differentiated Caco-2 nuclear extract, 2.5 fmol radioactive labeled probe (leftover seen in bottom of image), and 0.5 µg unspecific competitor poly(dI-dC). In addition, the wells contain: 1) Empty, 2) Unspecific competitor, 3) Non-labeled competitor probe, 4) Mutated non-labeled competitor probe, 5) Specific antibody (CDX2/HNF4α), 6) Unspecific antibody (HA). Lane 1: Two radiolabeled probe-protein

complexes are seen. Lane 2: No competition from the unspecific control probe. Lane 3: Both complexes outcompeted by the unlabeled probe. Lane 4: No competition by the mutated probe. Lane 5: complex 2 supershifted and the low-intensity complex 1 cannot be determined if is supershifted. Lane 6: No supershift from the unspecific antibody. **C)** Similar results as in B, but only one complex appears. **D)** Similar results as in C, except a somewhat stronger competition of the mutated probe in lane 4.

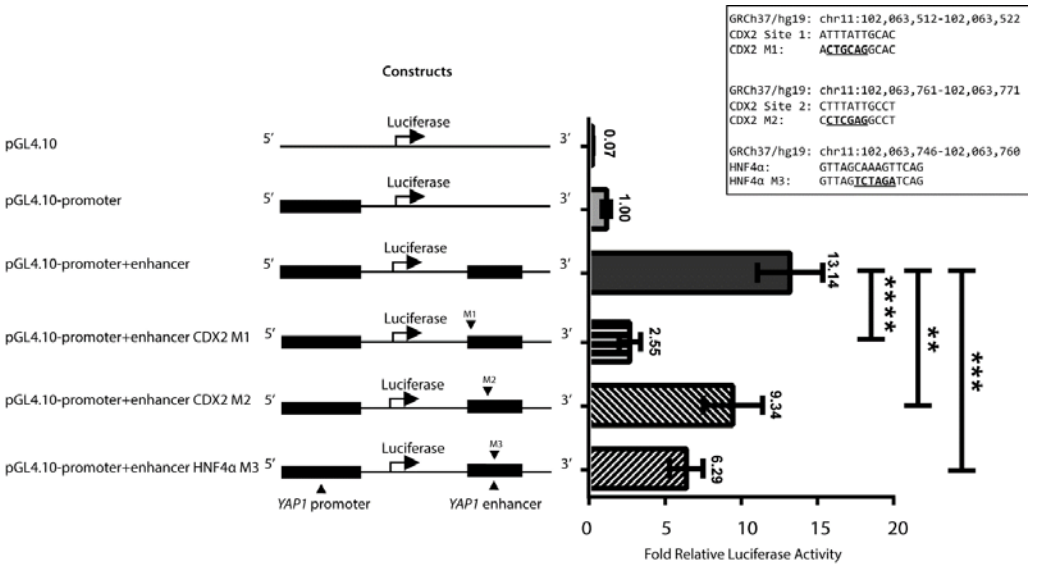


Figure 4 – pGL4.10 luciferase reporter constructs were generated by site-directed mutagenesis and assayed for relative activity in Caco-2 cells. The empty pGL4.10 control was digested with HindIII and the *YAP1* promoter was inserted, after which the *YAP1*-promoter construct was digested with SalI and the WT or mutated enhancer sequences inserted. The upper right corner contains the sequences of the three binding sites, CDX2-S1, CDX2-S2, and HNF4α, and their mutated sequence, M1, M2, and M3 with mutated nucleotides in bold. A double mutant was created with both the M2 and M3 mutations. The relative luciferase activity was assayed with a dual-light assay from Promega on a Glomax 96-well luminometer, using a pCMV-LacZ plasmid as a transfection control. All activity was normalized to the *YAP1*-promoter construct. The promoter activity is significantly higher than the background pgl4.10 activity, so the promoter is active in Caco-2 cells. The intragenic region acts as an enhancer with 13-fold upregulation from the promoter construct. The enhancer activity is decreased 5-fold by mutating the CDX2-S1, $P<0.0001$, 1.4-fold for the CDX2-M2 site $P<0.01$, and 2-fold for the HNF4α-M3 site.

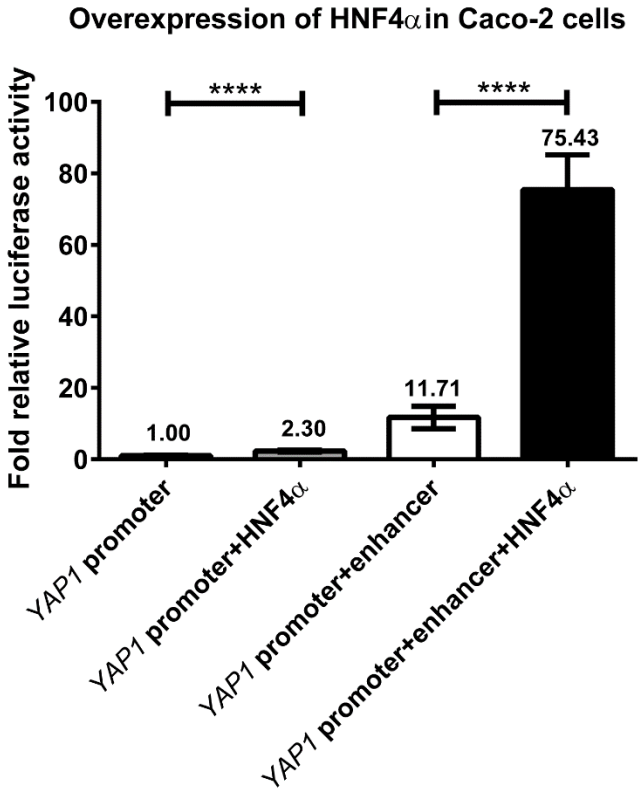


Figure 5 – Overexpression of HNF4α was achieved by co-transfecting *YAP1*-promoter and *YAP1*-promoter+enhancer constructs with CMV-HNF4α expression plasmids in Caco-2 cells. The intragenic region acts as an enhancer. HNF4α overexpression doubles the activity of the *YAP1*-promoter construct and increases the *YAP1*-promoter+enhancer constructs activity more than 6-fold. The fold relative luciferase activity is luciferase/beta-galactosidase activity normalized to a pGL-4.10 background and the *YAP1* promoter construct. Activity was measured using a Dual-Light assay by Promega. Bars are mean values, with error bars showing 95% CI. Data is statistically significant using two-tailed unpaired student's T-test with P-values under 0.0001 (***), n=4.

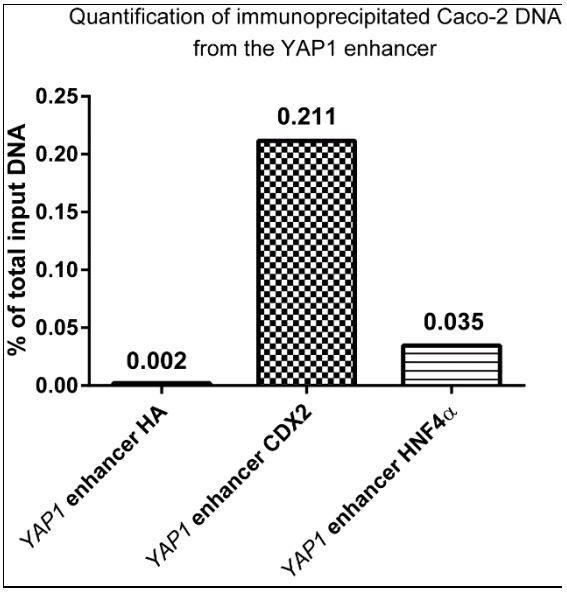


Figure 6 – PCR Quantification of Caco-2 DNA immunoprecipitated with either CDX2 or HNF4α antibodies. The amount of purified Caco-2 ChIP DNA was compared to the amount of input DNA in the *YAP1* enhancer, using specific primers (Table S2B). CDX2 enrichment was more than 100-fold compared to the HA control. HNF4a enrichment was more than 17-fold higher in the enhancer than in the control. Quantification was performed by RT-qPCR and values are means, n=2.

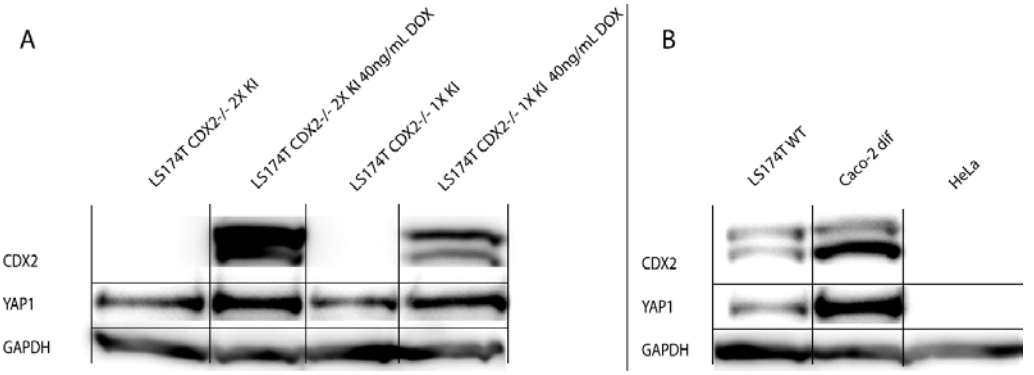


Figure 7 – Western Blot showing CDX2 and YAP1 protein expression levels with GAPDH as a reference. Cells were grown to confluence and proteins were extracted and run on a polyacrylamide gel before blotting on a PVDF membrane. Cells used: Caco-2, HeLa, LS174T WT, or LS174T cells with CDX2 knockout inducible by doxycycline; LS174T$^{CDX2-/-}$ 1X mono-allelic CDX2 Knock-In (CDX2 inserted in one allele), and LS174T$^{CDX2-/-}$ 2X bi-allelic CDX2 Knock-In (CDX2 inserted in both alleles). **A)** The

mono-and Bi-allelic CDX2 knock-in cells lines only express CDX2 when induced by Doxycycline and induction cause slightly increased YAP1 expression in the LS174$^{CDX2-/-}$ derived cell lines. Cells were induced by 40ng/mL dose added to the media for 24 hours. Induction of CDX2 increases YAP1 expression for the knock-in cell lines. **B)** Caco-2 and LS174T wildtype cells express CDX2 and YAP1, while HeLa does not. Horizontal lines denote where different blots were combined, and vertical lines denote where single wells from the same blot have been combined. Images were globally adjusted for contrast and brightness.

Supplementary tables

Reference genome is USCS GRCh37/hg19
YAP1 (NM_001130145.2) transcriptional start site: chr11:101,981,192
CDX2 and HNF4α sites found by in-silico JASPAR analysis in the *YAP1* promoter and enhancer

2 HNF4a and CDX2 sites were predicted in the *YAP1* promoter						
Model ID & name		Score	Relative score	Strand	Predicted site sequence	Loc Relative to TSS
MA0114.2 HNF4A		7.770	0.838	1	TCGGCCTTGGCCCTT	-451
MA0114.2 HNF4A		7.483	0.834	-1	GAGAACTTTTTCCCT	+111
MA0465.1 CDX2		6.636	0.852	1	GAACAAGAAAA	+34
MA0465.1 CDX2		7.394	0.864	1	AGGAAATAAAG	+44
2 HNF4a and **3** CDX2 sites were predicted in the *YAP1* enhancer						
Model ID	Site name	Score	Relative score	Strand	Predicted site sequence	Loc Relative to TSS
MA0114.2 HNF4A		10.168	0.870	1	TTTAACTTTGGTCCT	+82,340
MA0114.2 HNF4A	HNF4α	12.496	0.901	-1	CTGAACTTTGCTAAC	+82,555
MA0465.1 CDX2	CDX2-S2	12.227	0.936	-1	GTGCAATAAAT	+82,321
MA0465.1 CDX2	CDX2-S1	14.121	0.965	-1	AGGCAATAAAG	+82,570
MA0465.1 CDX2		10.212	0.906	1	CTCCCATAAAG	+82,626

Table S1 – Transcription factor binding sites from an *in-silico* analysis using the online JASPAR tool (URL: http://jaspar.genereg.net), (Mathelier et al., 2016). *YAP1* input sequence is from the USCS GRCh37/hg19 genome assembly. The *YAP1* promoter input sequence spans 972bp at chr11:101,980,493-101,981,464) and the *YAP1* enhancer input sequence spans 599bp at (chr11:102,063,361-102,063,959). Two CDX2 and two HNF4α sites were predicted in the promoter region, using the MA0465.1 and MA0114.2 matrices and a relative score cutoff of 80% and 85%, respectively. Two CDX2 and three HNF4α sites were predicted in the enhancer with a relative score cutoff of 85%. The location of the binding sites in the genome is given relative to the transcription start site (TSS) of *YAP1* at chr11:101,981,192.

A

Primers used for reporter constructs		
Primer name	Primer Sequence	Restriction site
YAP1 Prom-F	CCGTTTACCCCTCTCAAGTG	
YAP1 Prom-R	GCTGTCCTCGCTCTCAGG	
Inf-*YAP1* Prom-F	CTCGGCGGCC**AAGCTT**CCGTTTACCCCTCTCAAGTG	HindIII
Inf-*YAP1* Prom-R	CCGGATTGCC**AAGCTT**GCTGTCCTCGCTCTCAGG	HindIII
YAP1 enhancer-F	TGACTGGATTAGACTGGATGCT	
YAP1 enhancer-R	GGAAAAAGAAAATGTAGTGAGAGC	
Inf-*YAP1* enhancer-F	ATAAGGATCC**GTCGAC**TGACTGGATTAGACTGGATGCT	Sal1
Inf-*YAP1* enhancer-R	AAGGGCATCG**GTCGAC**GGAAAAAGAAAATGTAGTGAGAGC	Sall

B

Primers used for ChIP analysis	
Primer name	Primer Sequence
YAP1-enhancer F	CCTCTGGTCACAGTGTGGAA
YAP1-enhancer R	TTGCCTGGACATACTCACCA

Table S2 - A) Primers used to produce the luciferase reporter constructs. Primers with In-Fusion tails contain the prefix “inf” and the tail sequences are underlined. Bold text denotes the sequence of restriction sites replacing transcription factor binding sites via site-directed mutagenesis. **B)** Primers used for real-time PCR in the analysis of chromatin immune precipitated DNA from Caco-2 nuclear extract.

Manuscript II

2.2 Paper II CDX2 regulates interleukin-33 gene expression in intestinal epithelial cell lines

Sylvester Larsen[1,2], Jakob Benedict Seidelin[3,] Johanne Davidsen[1,4], Katja Dahlgaard[1], Claus Henrik Nielsen[5], Eric Paul Bennett[6], Ole Birger Pedersen[2], Mehmet Coskun[7,3#], and Jesper Thorvald Troelsen[1*#]

[1]Department of Science and Environment, Roskilde University, Roskilde, Denmark. [2]Department of Clinical Immunology, Næstved Hospital, Næstved, Denmark. [3]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Herlev, Denmark. [4]Department of Surgical Gastroenterology, Zealand University Hospital, Køge, Denmark. [5]Institute for Inflammation Research, Center for Rheumatology and Spine Diseases, section 7521, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark. [6]Copenhagen Center for Glycomics, Department of Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [7]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark.

[#] Shared last authors, Corresponding Author*

Manuscript is submitted to the journal Cellular and Molecular Gastroenterology and Hepatology

The aim of the study:

In paper II, the aim was to investigate the role of CDX2 in regulating *IL33* gene expression in colon cell lines. It is known that dysregulation of *IL-33* can lead to intestinal inflammation and a variety of other pathologies, but little is known about its transcriptional regulation and specific role in the intestinal homeostasis. Since Chip-Seq data suggested that the *IL33* gene contained binding sites for the major intestinal transcription factor CDX2, it was attempted to clarify the role of CDX2 in regulating *IL33* expression in intestinal cells. This was done by analyzing the chromatin landscape of the *IL33* gene to determine the position of CDX2 binding sites in the promoter and an intragenic enhancer. Subsequently, the CDX2 inducible colon cancer cell line developed

in paper III was sued to investigate the ability of CDX2 to affect *IL33* gene expression through binding to its enhancer.

Key findings:

- It was revealed that *IL33* is a potential target gene of the intestinal transcription factor CDX2 in colon cancer cell lines using bioinformatic analysis of the chromatin landscape of *IL33*.
- The binding of CDX2 to the *IL33* enhancer was verified by using chromatin immunoprecipitation in LS174T colon cells.
- Quantitative PCR analysis showed that *IL33* mRNA was especially high in primary colonic cells as well as in LS174T colon cancer cells, compared to several other colon cancer cell lines.
- Induction of CDX2 expression in the LS174T colon cell model developed in paper III demonstrated that *IL33* mRNA expression is significantly upregulated by CDX2
- CDX2 was identified as an important transcription factor for *IL33* promoter activity with importance for high expression levels in colonic epithelial cells.

Title: CDX2 regulates interleukin-33 gene expression in intestinal epithelial cell lines

Short title: CDX2 regulates IL-33 in intestinal cells

Authors: Sylvester Larsen[1,2], Jakob Benedict Seidelin[3,] Johanne Davidsen[1,4], Katja Dahlgaard[1], Claus Henrik Nielsen[5], Eric Paul Bennett[6], Ole Birger Pedersen[2], Mehmet Coskun[7,3#], and Jesper Thorvald Troelsen[1*#]

Affiliations: [1]Department of Science and Environment, Roskilde University, Roskilde, Denmark. [2]Department of Clinical Immunology, Næstved Hospital, Næstved, Denmark. [3]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Herlev, Denmark. [4]Department of Surgical Gastroenterology, Zealand University Hospital, Køge, Denmark. [5]Institute for Inflammation Research, Center for Rheumatology and Spine Diseases, section 7521, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark. [6]Copenhagen Center for Glycomics, Department of Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [7]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark.

[#]Shared last authors

* To whom correspondence should be addressed: Jesper T. Troelsen: Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark; troelsen@ruc.dk; Tel.(+45) 4674-2728

Grant Support: This study was funded by the Independent Research Fund Denmark (grant ref 4004-00140B), the Lundbeck Foundation, Fonden til Lægevidenskabens

Fremme (the A. P. Møller Foundation), the Foundation of Aase and Ejnar Danielsen, Læge Sofus Carl Emil Friis & Hustru Olga Doris Friis' Legat, the region zealand foundation for health and research, and the Danish National Research Foundation [DNRF107]. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest: The authors disclose no conflicts.

Abstract

Interleukin 33 (IL-33) is a pleiotropic member of the IL-1 family of cytokines. Dysregulation of IL-33 has been implicated in the pathogenesis of several autoimmune and inflammatory diseases. Despite the fact that IL-33 has been detected in multiple tissues and cell types, limited knowledge exists about the transcriptional regulation of *IL33*.

qPCR assays were used to measure *IL33* mRNA expression in primary colonic epithelial cells from healthy controls and several colon cancer cell lines. This revealed particularly high expression levels in primary colonic cells and LS174T cells.

Bioinformatic analysis revealed that *IL33* is a potential target gene of the intestinal transcription factor CDX2 in LS174T intestinal cells. CDX2 was found to bind to a potential intronic enhancer in the *IL33* gene. Transfection of *IL33* promoter reporter constructs containing the intronic enhancer region showed that it is a potent enhancer of *IL33* promoter activity in LS174T cells. Furthermore, chromatin immunoprecipitation experiments verified that CDX2 binds to the *IL33* enhancer in LS174T cells. Induction of CDX2 expression in LS174T cells demonstrated that *IL33* mRNA expression is significantly upregulated by CDX2.

In conclusion, we identified CDX2 as an important transcription factor for *IL33* promoter activity with importance for high expression levels in intestinal epithelial cells.

Background & Aims

Interleukin (IL)-33 is a member of the IL-1 family of cytokines and is constitutively expressed by a diverse range of cells including epithelial cells, endothelial cells, macrophages, dendritic cells, keratinocytes, fibroblasts, fibrocytes, and smooth muscle cells, with especially high expression detected in the small intestine, stomach, lung, spinal cord, brain, skin, tonsils, and lymph nodes [1-5]. Nevertheless, how *IL33* gene expression is controlled is still largely unknown. Given the clear involvement of IL-33 in homeostasis and diseases, a better understanding of *IL33* gene regulation is crucial.

IL-33 is involved in tissue and metabolic homeostasis as well as infectious and inflammatory diseases including rheumatoid arthritis, allergies, intestinal inflammation, cardiovascular diseases, and disorders of the central nervous system [6-11]. Furthermore, preliminary data shows that IL-33 plays a role in skin wound healing [12, 13]. Furthermore, it was discovered that increased IL-33 expression leads to accelerated cellular transformation, suppression of innate antiviral immunity and dysregulation of Th1, Th2, and Treg cells [14-16].

IL-33 is synthesized as a full-length 30kDa bioactive protein, but it can be cleaved or inactivated by caspases or proteases to modify its function [17, 18]. During inflammation, the bioactivity of IL-33 can be increased up to 30-fold in inflamed tissues by inflammatory proteases that can partially proteolyse the full-length bioactive IL-33 into N-terminally truncated shorter mature forms (18-22kDa) [19-21]. Mature 18kDa recombinant IL-33 has been shown to promote colon cancer cell stemness via JNK activation and macrophage recruitment [22].

The full-length bioactive IL-33 has also shown to be released by necrotic or damaged cells in response to stress conditions and necrosis caused by injury, infections, or inflammation, thereby acting as an "alarmin" that amplifies immune responses [23, 24]. The full-length peptide signals through a heterodimeric receptor complex comprising the transmembrane receptor suppressor of tumorigenicity 2 (ST2) and the co-receptor IL-1 receptor accessory protein (IL-1RAcP) [25]. In cells which express ST2, IL-33 interacts with ST2 to activate the nuclear factor (NF)-κB and mitogen-activated protein kinase (MAPK) signaling pathways [3].

Intracellular full-length IL-33 may function as a nuclear factor with transcriptional activity. It can bind to chromatin and acts as a negative regulator of transcription through a chromatin-binding motif [2, 26]. It also binds to the p65 subunit of NF-κB, which reduces its binding to NF-κB response elements, and reduces p65-induced transcriptional activation [27]. However, IL-33 also functions as a transcriptional activator by upregulating the expression of NF-κB, interleukin 6 (IL-6) and ST-2 in endothelial cells [28, 29].

In human endothelial cells *IL33* expression is regulated by the cytokine tumor necrosis factor alpha (TNF), vascular endothelial growth factor (VEGF), and NOTCH signaling [4, 30]. Additionally, a study investigated *IL33* promoter activity in human epithelial cell lines and showed that interferon gamma is an activator of promoter activity [31]. Moreover, several studies have reported that IL-33 protein expression is stimulated by growth factors, transcription factors, cytokines, and other proteins in both murine and human cells [30, 32-35].

In the intestines, gene regulation is controlled by a transcription factor network of which *Caudal*-related homeobox transcription factor 2 (CDX2) is a key component.

Because of the major role of CDX2 in regulating intestinal genes, we investigated whether CDX2 is involved in the regulation of *IL33* gene expression in the intestinal epithelium. CDX2 is specifically expressed in the adult intestine and it is essential in intestinal homeostasis and control of the balance between cell proliferation and differentiation [36, 37]. Moreover, CDX2 is a crucial factor in the regulation of genes related to epithelial functions and intestinal cell fate. Several studies have demonstrated that CDX2 is directly involved in the regulation of genes including transcription factors, digestive enzymes, nutrient transporters, surface proteins, and genes in the Wnt/β-catenin degradation complex with essential functions in the intestinal epithelium [38-45].

Methods

Cultivation of cell lines and primary colonic epithelial cells

Human colon cancer cell lines Caco-2, HT-29, LS174T, SW480 and DLD-1 (American Type Culture Collection, Rockville, MD, USA) and CDX2-inducible LS174T$^{CDX2-/-}$ cells [46] were cultured and grown as monolayers in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% glutamine. All cell lines were maintained at 37°C and 5% CO_2. Primary colonic epithelial cells were isolated from a pool of six biopsies obtained from healthy control subjects (i.e., patients undergoing a routine endoscopy where all investigations subsequently turned out to be normal) as described previously [47]. Biopsies were collected at Dept. of Gastroenterology, Herlev Hospital, Denmark from healthy controls given their written informed consent.

CDX2 knock-down and induction with doxycycline

The CDX2-inducible knockout cells (LS174T$^{CDX2-/-}$) are described in [46]. In short, zinc-finger nucleases were utilized to generate LS174T cells with a bi-allelic *CDX2* knockout. Subsequently, a cassette containing the *CDX2* gene under the control of a doxycycline activated Tet-3G induction system was inserted into the safe-harbor locus AAVS1 of one allele. Using this system it is possible to control the expression of *CDX2* by stimulation with doxycycline without any leakiness [46].

IL33 qRT-PCR and Luminex analysis

The NucleoSpin RNA kit (Macherey-Nagel, Düren, Germany) was used for extraction of total RNA from Caco-2, HT29, LS174T, LS174T$^{CDX2-/-}$, SW480, DLD1, and from isolated colonic epithelial cells. 200 ng of each RNA sample was used for cDNA synthesis with SuperScript III Reverse Transcriptase (Invitrogen, Paisley, UK). Quantitative RT-PCR reactions were performed on a LightCycler 480 instrument (Roche, Mannheim, Germany) using the Maxima SYBR Green qPCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturer's instructions. All experiments were normalized to the housekeeping gene Ribosomal Protein Large P0 (RPLP0).

LS174T^{WT} and CDX2-inducible LS174T$^{CDX2-/-}$ cells were grown in DMEM supplemented with 10% FBS and 1% glutamine. CDX2 expression was induced with 0, 2, 4, 6, 8, 10, 12, or 14ng/mL doxycycline for 24 hours; all samples were in quadruplicates. The supernatant was stored at -80°C for about a month until measured on a Bio-plex 2200 system (Bio-Rad, Hercules, CA, USA). IL-33 protein was measured on undiluted supernatants using the Bio-Plex Pro Human Th17 assay for IL-33 (171-BA012M) according to manufacturer protocol.

Construction of luciferase reporter constructs

The *IL33* promoter and enhancer sequence were PCR amplified and gel purified. Amplification was done with non-infusion primers and subsequently extended by PCR with In-Fusion primers. Primers were ordered from Eurofins genomics (Ebersberg, Germany) (supplementary table S1A). pGL4.10 vector (Promega, Madison, WI, USA) was digested with *Hin*dIII and gel purified. The pGL4.10-*IL33*-promoter plasmid was

constructed by In-Fusion cloning (Clontech, Fremont, CA, USA) purified promoter sequence and digested vector in 1:2 ratio, according to manufacturer's protocol. 2.5µl cloning reaction was used for transformation of One Shot TOP10 chemically competent E. *coli* (Thermo Fisher Scientific, Waltham, MA, USA). Colonies were isolated and Sanger sequenced (Beckman Coulter Genomics, United Kingdom). For the construction of the pGL4.10-*IL33*-promoter+enhancer plasmid, the pGL4.10-*IL33*-promoter plasmid was digested with *Sal*I, gel purified, and In-Fusion cloned with enhancer insert in a 1:2 molar ratio.

Measuring luciferase activity of reporter constructs

Measurements were carried out on a GloMax® 96 Microplate Luminometer using the Dual-Light™ Luciferase & β-Galactosidase Reporter Gene Assay System (Thermo Fisher Scientific, Waltham, MA, USA). Cells were rinsed three times with 1x PBS and lysed with 130µL TROPIX lysis solution containing 0.5 mM DTT and incubated for 10 min on ice. 10µL of lysate was transferred to a 96 well GloMax luminometer light plate, and luciferase activity was measured using 5 sec integration time and 2 sec delay. Beta-galactosidase activity was measured 45 minutes after luciferase activity.

Chromatin immunoprecipitation assays

CDX2 wild-type (LS174T^{WT}) cells grown to five days after confluence in a 30x30 cm culture dish were cross-linked and sonicated as described previously to generate fragments of ~0.2 to 1.2 kb [48]. Briefly, immunoprecipitation (IP) was done in four replicates and performed overnight at 4ºC with an antibody specific for either human

CDX2, (α-CDX2 clone CDX2-88; #MU392A-UC, BioGenex, Fremont, CA, USA) or as a negative control an antibody specific for the influenza hemagglutinin (HA) epitope (rabbit polyclonal α-HA, Y-11 X, #SC-805 X; Santa Cruz Biotechnology Inc, Heidelberg, Germany). Immunocomplexes were recovered with 50 µl protein A/G beads (Invitrogen, Grand Island, NY, USA). Real-time PCR was carried out using DNA immunoprecipitated with CDX2 and HA or on non-IP material from the LS174T^{WT} cell line (Input; representing 1% of the total amount used in IP) using primers designed for the peak in the *IL33* region (Figure 2A). The primers were obtained from TAG Copenhagen A/S, Denmark (supplementary table S1B).

Protein extraction and immunoblotting

LS174T^{WT} and CDX2-inducible LS174T$^{CDX2-/-}$ cells were seeded in 6-well plates at 300,000 cells/well. After 48hrs, the media was changed to either doxycycline free media or media with varying concentrations of doxycycline. At 72hrs, cells were rinsed with cold PBS and lysed in RP1 lysis buffer (Macherey-Nagel, Düren, Germany) and proteins were purified according to the manufacturer's protocol. Protein concentration was determined by Bradford analysis (Bio-Rad, Hercules, CA, USA). 10µg protein was mixed 1:4 (v/v) with Bolt loading buffer and 1:10 (v/v) with Bolt sample reducing agent (Thermo Fisher Scientific, Waltham, MA, USA). Samples were incubated at 70°C for 10min and loaded on a Bolt 4-12% Bis-Tris Plus gel along with prestained protein marker PageRuler both from (Thermo Fisher Scientific, Waltham, MA, USA). SDS-PAGE was performed in 1X Bolt MOPS running buffer, (Thermo Fisher Scientific, Waltham, MA, USA) for 50 min at 150V. Gels were transferred by wet-electrotransfer

to PVDF membranes for 75 min at 25 V and 100mA in 1X NuPage transfer buffer, (Thermo Fisher Scientific, Waltham, MA, USA). Membranes were blocked with dry skim milk diluted to 5% in wash buffer (1X PBS with 0.1% Tween-20) for 1 hour at room temperature, washed with Wash buffer 3 x 7min and incubated ON at 4°C with primary antibody diluted in Dilution buffer (2.5% skim milk in wash buffer). Membranes were washed 3 x 7min and incubated with secondary antibody for 1 hour at RT and washed 3 x 7min. Bands were visualized by incubating with the ECL solution SuperSignal™ West Dura Extended Duration Substrate (Thermo Fisher Scientific, Waltham, MA, USA). Primary antibodies were mouse monoclonal anti-CDX2 (1:1000; #MU392A-UC, BioGenex, Fremont, USA) and anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (1:20.000; #10R-G109a, Fitzgerald, Acton, MA, USA). The secondary antibody was a horseradish peroxidase-labelled goat anti-mouse antibody (1:10000; #32230, Thermo Fisher Scientific, Waltham, MA, USA).

Statistical analysis

The analysis was carried out using two-tailed Student's t-test or one-way ANOVA with multiple comparisons using Dunnett's test or Tukey's test. P-values <0.05 were considered significant. Values are presented as means with SD or 95% CI. Significance levels are given as * $P<0.05$, ** $P<0.01$, *** $P<0.001$, **** $P<0.0001$.

Results

IL33 mRNA expression by intestinal epithelial cells

The expression of *IL33* transcripts was measured by qRT-PCR in several well-characterized intestinal epithelial cell lines as well as in primary colonic epithelial cells isolated from healthy subjects. The highest expression level of *IL33* was found in primary colonic epithelial cells, closely followed by LS174T^{wt} and HT29 cells while the *IL33* expression was either undetected or very low in SW480, DLD-1, and Caco-2 cell lines (Figure 1).

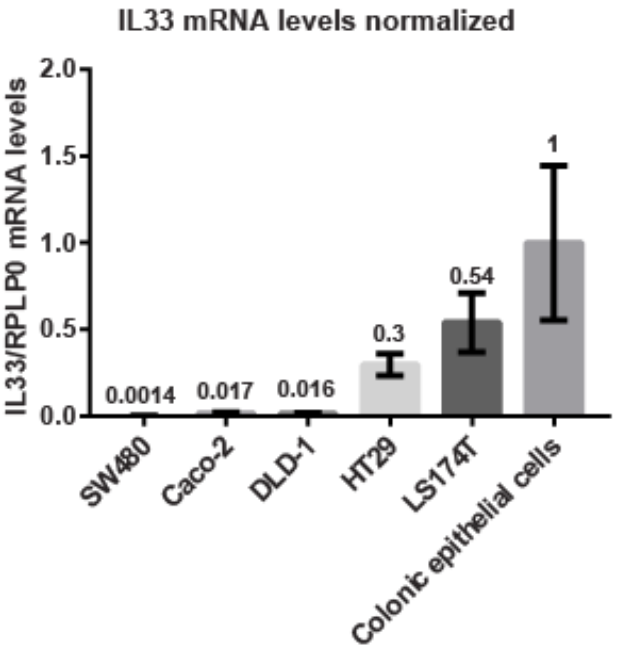


Figure 1. Expression of *IL33* mRNA in intestinal cell lines and primary human colonic epithelial cells. Total RNA from SW480, Caco-2, DLD-1, HT29, LS174T cells, and colonic epithelial cells isolated from healthy humans were extracted, converted to cDNA, and measured by q-PCR. Data were normalized to the reference gene RPLP0. Data are represented as mean values with SD error bars (*n=4*).

CDX2 binds to *IL33* regulatory regions in LS174T cells

Recently, a genome-wide CDX2 chromatin immunoprecipitation (ChIP) sequence analysis (ChIP-seq) of the human colon cancer cell line LS174T was published [46]. By analysis of these data, we identified a clear CDX2 ChIP-seq signal within the *IL33* gene locus (Figure 2A). The ChIP-seq peak mapped to the *IL33* gene in GRCh37/hg19 at the location chr9:6243827-6244384. Sequence analysis identified six potential CDX2-binding sites within the *IL33* ChIP-seq peak, indicating a potential regulatory region (Figure 2B). CDX2 ChIP on LS174T^{WT} cells was performed to validate the ChIP seq findings and to determine if *IL33* is a direct CDX2 target. Using primers specific to the potential CDX2 binding region, we found that the amount of CDX2-IP DNA was about 9-fold higher than that of the negative control HA-IP DNA (p<0.001) (Figure 2C). This shows that CDX2 is associated with the region and is a potential regulator of the *IL33* gene expression in LS174T^{wt} intestinal cells.

Taken together with the ChIP-seq data, the *IL33* expression profiles indicated that LS174T cells were a suited model to investigate the intestinal transcriptional regulation of *IL33* by CDX2. Thus, all subsequent analyses were performed using LS174T cells.

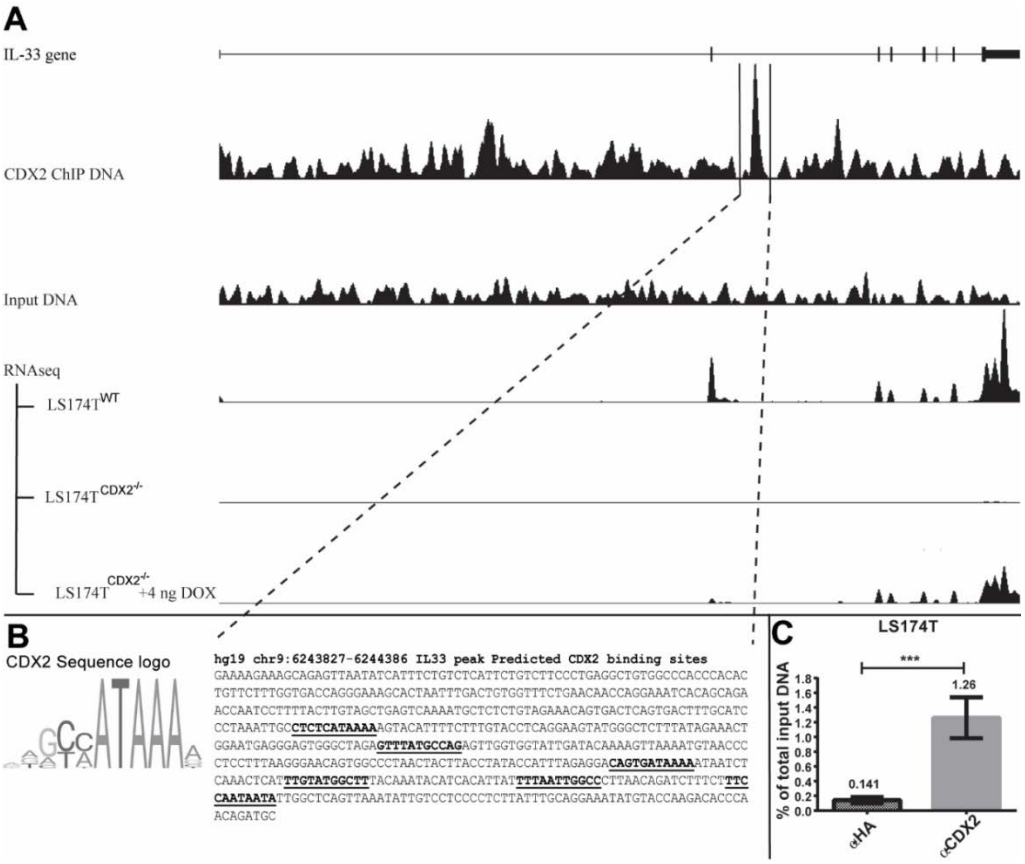


Figure 2. *IL33* is a potential CDX2 target in LS174T cells. (**A**) View of the 42,198bp *IL33* gene (uc003zjt.3, NM_033439) from the UCSC Genome Browser GRCh37/hg19. Imported data tracks include CDX2 ChIP-seq reads of both immunoprecipitated and input DNA (representing 1% of the total amount used in IP) signal, shown as density graphs [46]. RNA-Seq data from deep sequencing transcriptomic analysis from LS174T^{WT} and CDX2-inducible LS174T$^{CDX2-/-}$ are shown with or without induction with 4 ng/mL doxycycline for 24 hours [46]. Data are displayed as density graphs. (**B**) The position-weight matrix of the CDX2-binding motif [ID: MA0465.1] from the JASPAR database (http://jaspar.genereg.net/) [49] including six (bold and underlined) predicted CDX2 target sites in the sequence of the major CDX2 binding element at the marked area (chr9:6243827-6244384) in the *IL33* gene. (**C**) qPCR of CDX2 and hemagglutinin (negative

control) chromatin immunoprecipitated material from LS174T^{WT} cells generated with primer pairs located within the CDX2 ChIP-seq peak in the *IL33* locus (showed in 2A). The relative enrichment of CDX2 binding is shown as a percentage of total input DNA and is represented as means with SD error bars (n=4), *** P<0.001.

CDX2 regulates expression of *IL33* RNA in LS174T cells

An LS174T derived cell-line was used to determine to which extent CDX2 regulates *IL33* expression in intestinal cells. The CDX2-inducible LS174T$^{CDX2-/-}$ cells express CDX2 only in response to doxycycline stimulation [46]. CDX2 target genes in LS174T cells were identified by RNA-seq by comparing LS174T^{WT}, non-stimulated LS174T$^{CDX2-/-}$, and doxycycline-stimulated LS174T$^{CDX2-/-}$ cells [46].

The *IL33* expression profile from these RNA-seq data revealed that in knockout cells with no CDX2 expression, the *IL33* expression was very low, whereas the *IL33* mRNA level was clearly increased in LS174T$^{CDX2-/-}$ cells treated with doxycycline for 24 hours (Figure 3, upper panel). This indicates that CDX2 is an important activator of *IL33* expression in LS174T cells.

Subsequently, CDX2-inducible LS174T$^{CDX2-/-}$ cells were treated with increasing concentrations of doxycycline (0–10 ng/mL). Western blotting analysis of CDX2 showed an increased level of CDX2 protein following the doxycycline induction (Figure 3, lower panel). In addition, a clear doxycycline concentration-dependent increase of *IL33* mRNA expression was observed (Figure 3, upper panel). This suggests that induction of CDX2 upregulates *IL33* mRNA expression.

Luminex analysis showed that supernatants from LS174T$^{CDX2-/-}$ cell induced with 0-14ng/mL doxycycline contained no detectable level of IL-33 protein (Data not shown).

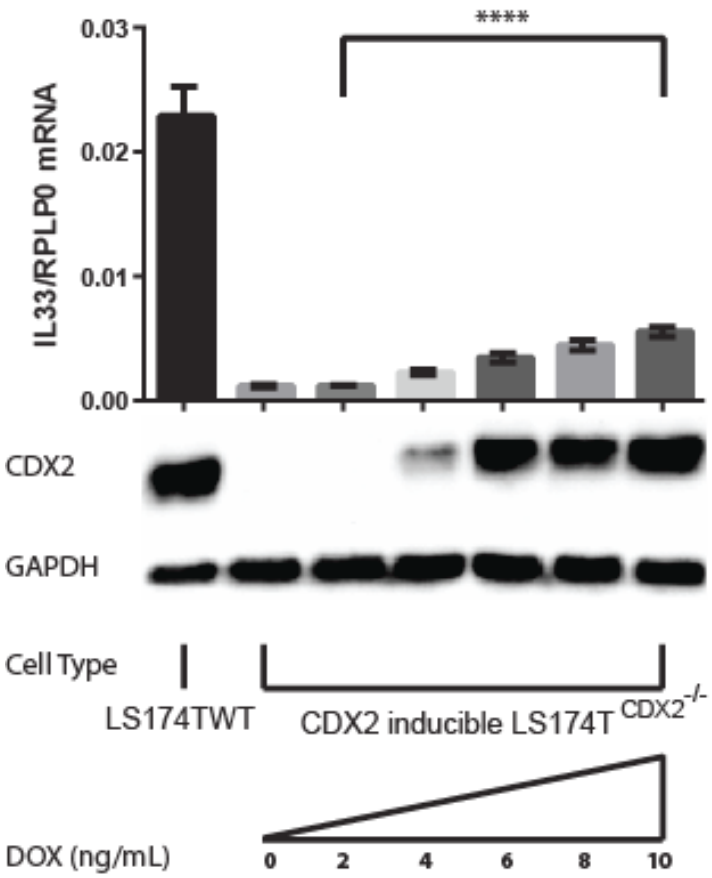


Figure 3. (Top) Quantitative RT-PCR of *IL33* mRNA expression in LS174T^{WT} and CDX2-inducible LS174T$^{CDX2-/-}$ cells exposed to increasing concentrations of doxycycline, normalized to the reference gene RPLP0. CDX2 expression was induced by incubation with 4 ng/mL or 10 ng/mL doxycycline for 24 hours as indicated. Data are represented as mean values with SD error bars, (n=4), **** denotes P<0.0001 using one-way ANOVA with multiple comparisons. (**Bottom**) Western blot of CDX2 protein levels in LS174T^{WT} and CDX2-inducible LS174T$^{CDX2-/-}$ cells after exposure to different concentrations of doxycycline for 24 hours. GAPDH was used as loading control.

The activity of the *IL33* regulatory elements

To investigate the regulatory effect of CDX2 on *IL33* expression, we used promoter reporter assay analysis. We constructed luciferase reporter plasmids containing the *IL33* promoter with or without the potential enhancer region identified in the ChIP-seq analysis (Figure 2A) and transfected them into LS174T^{WT} and CDX2-inducible LS174T$^{CDX2-/-}$ cells. This showed that the intragenic enhancer region increased the *IL33* promoter activity more than 45-fold from the promoter level (Figure 4), which demonstrates that the regulatory element is a powerful enhancer of *IL33* in LS174T cells.

To investigate the role of CDX2 for the enhancer effect, CDX2-inducible LS174T$^{CDX2-/-}$ cells were transfected with the *IL33* promoter and promoter/enhancer constructs and CDX2 expression was induced with varying amounts of doxycycline stimulation. Transfection of the *IL33* promoter/enhancer construct into non-induced cells, that do not express any CDX2, increased the reporter gene expression 8-fold from the pGL4.10-*IL33*-promoter plasmid alone, suggesting that the enhancer activity is also regulated by other factors than CDX2 (Figure 4). However, when CDX2 was induced with 4 ng/mL or 6 ng/mL doxycycline for 24 hours, the effect of the enhancer increased 19.3-fold and 31.7-fold, respectively, over the non-induced pGL4.10-*IL33*-promoter plasmid, demonstrating a CDX2 dependent upregulation of the *IL33* promoter activity (Figure 4).

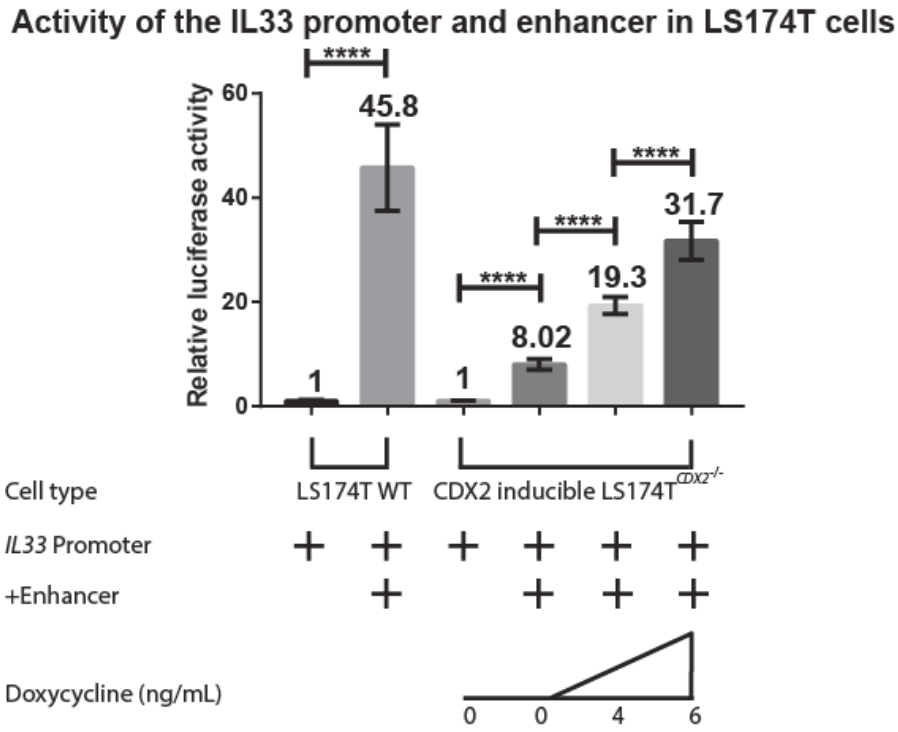


Figure 4. The activity of the *IL33* promoter and enhancer in LS174T cells. pGL4.10 Luciferase reporter constructs containing *IL33*-promoter and enhancer elements were transfected into LS174T^{wt} and CDX2-inducible LS174T$^{CDX2-/-}$ cells. Cells were exposed to different concentrations of doxycycline for 24 hours as indicated. Their relative luciferase/beta-galactosidase activity was measured in cell lysates using a dual light assay. Activity was normalized to the constructs with only the *IL33* promoter. Bars represent mean values with SD error bars using one-way ANOVA with multiple comparisons or two-tailed Student's t-test, **** denotes P>0.0001, n=4)

Discussion

The roles of IL-33 in homeostasis and in infectious and inflammatory diseases are well documented [6, 7, 9]. However, the transcriptional regulation of *IL33* gene expression remains obscure. Revealing the regulatory mechanisms behind of *IL33* expression may enhance our understanding of the role of IL-33 in the intestinal tissue.

In silico transcription factor binding site analysis of the intestinal cell line LS174T identified a clear peak in intron 2 of *IL33* that was determined to contain several CDX2 binding sites, and promoter-reporter assays in CDX2 inducible LS174T$^{CDX2-/-}$ cells revealed that this region may be a potent enhancer of *IL33* promoter activity.

IL33 mRNA levels varied greatly between the intestinal cell lines investigated and were especially high in colonic epithelial cells isolated from healthy patients., as shown by qPCR (Figure 1). While HT-29 and SW480 express relatively little CDX2, the other cell lines express comparable CDX2 levels [42, 46, 50, 51], suggesting that expression is dependent on other factors than CDX2.

However, the *IL33* expression profile from the RNA-seq data of CDX2 knockout and CDX2 re-induced LS174T cells suggested that CDX2 is important for *IL33* expression (Figure 2A). This was further supported by the up-regulation of *IL33* mRNA expression by CDX2 protein (Figure 3). However, the transfections with promoter reporter constructs showed that the construct containing the *IL33* enhancer was still able to activate the *IL33* promoter in cells without CDX2 induction albeit at a lower level (Figure 4). This further supports that other factors than CDX2 may act as activators on the *IL33* enhancer to uphold a low basal level of *IL33* expression. It is well established

that the regulation of CDX2 target genes is also regulated by other members of the intestinal transcription factor network, such as hepatocyte nuclear factor 4 alpha (HNF4α) and transcription factor 7-like 2 (TCF7L2) [41, 48, 52-55].

Since inflammatory bowel diseases and colorectal cancers have been linked to the dysregulation of CDX2 [50, 56-58] and dysregulation of IL-33 has been linked to inflammatory bowel disease and colorectal cancers [10, 59, 60], we hypothesize that *IL33* expression might be affected by inflammatory bowel diseases and colorectal cancers through dysregulation of CDX2. Under normal physiological conditions, on the other hand, we suggest that IL-33 is constitutively expressed in intestinal epithelial cells under the influence of normal CDX2 expression. The intracellular IL-33 pool may simply be as a store of alarmins that can be released during damage of the intestinal barrier [24], or it may play a homeostatic role as transcription factor, by suppressing NF-κB signaling and production of pro-inflammatory cytokines like IL-6 [27, 28]. Furthermore, secreted IL-33 may increase mucus production by activating Muc-gene transcription in an autocrine or paracrine manner [61].

This study provides the first evidence that the intestinal specific transcription factor CDX2 regulates the expression of *IL33* in intestinal cell lines. Furthermore, we identified an intronic enhancer of *IL33* gene expression in LS174T cells. We suggest that CDX2 adds an intestinal-specific layer of control to the expression of IL-33 through transcriptional regulation of *IL33*.

REFERENCES

1. Baekkevold ES, Roussigne M, Yamanaka T, et al. Molecular characterization of NF-HEV, a nuclear factor preferentially expressed in human high endothelial venules. Am J Pathol 2003;163:69-79.
2. Carriere V, Roussel L, Ortega N, et al. IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo. Proc Natl Acad Sci U S A 2007;104:282-7.
3. Schmitz J, Owyang A, Oldham E, et al. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. Immunity 2005;23:479-90.
4. Kuchler AM, Pollheimer J, Balogh J, et al. Nuclear interleukin-33 is generally expressed in resting endothelium but rapidly lost upon angiogenic or proinflammatory activation. Am J Pathol 2008;173:1229-42.
5. Moussion C, Ortega N, Girard JP. The IL-1-like cytokine IL-33 is constitutively expressed in the nucleus of endothelial cells and epithelial cells in vivo: a novel 'alarmin'? PLoS One 2008;3:e3331.
6. Liew FY, Girard JP, Turnquist HR. Interleukin-33 in health and disease. Nat Rev Immunol 2016;16:676-689.
7. Liew FY, Pitman NI, McInnes IB. Disease-associated functions of IL-33: the new kid in the IL-1 family. Nat Rev Immunol 2010;10:103-10.
8. Palmer G, Gabay C. Interleukin-33 biology with potential insights into human diseases. Nat Rev Rheumatol 2011;7:321-9.
9. Molofsky AB, Savage AK, Locksley RM. Interleukin-33 in Tissue Homeostasis, Injury, and Inflammation. Immunity 2015;42:1005-19.
10. Seidelin JB, Coskun M, Kvist PH, et al. IL-33 promotes GATA-3 polarization of gut-derived T cells in experimental and ulcerative colitis. J Gastroenterol 2015;50:180-90.
11. Arshad MI, Khan HA, Noel G, et al. Potential Therapeutic Aspects of Alarmin Cytokine Interleukin 33 or Its Inhibitors in Various Diseases. Clin Ther 2016;38:1000-1016 e1.
12. Ansell DM, Hardman MJ. Do not be alarmed: understanding IL33-ST2 signalling in wound repair. Exp Dermatol 2016;25:22-3.
13. Lee JS, Seppanen E, Patel J, et al. ST2 receptor invalidation maintains wound inflammation, delays healing and increases fibrosis. Exp Dermatol 2016;25:71-4.
14. Ohta S, Tago K, Funakoshi-Tago M, et al. Intracellular NF-HEV/IL-33 harbors essential roles in Ras-induced cellular transformation by contributing to cyclin D1 protein synthesis. Cell Signal 2016;28:1025-36.
15. Matsuda A. A novel function of IL-33: suppression of innate antiviral immunity. J Thorac Dis 2016;8:E1364-E1366.

16. Braun H, Afonina IS, Mueller C, et al. Dichotomous function of IL-33 in health and disease: From biology to clinical implications. Biochem Pharmacol 2018;148:238-252.
17. Talabot-Ayer D, Lamacchia C, Gabay C, et al. Interleukin-33 is biologically active independently of caspase-1 cleavage. J Biol Chem 2009;284:19420-6.
18. Afonina IS, Muller C, Martin SJ, et al. Proteolytic Processing of Interleukin-1 Family Cytokines: Variations on a Common Theme. Immunity 2015;42:991-1004.
19. Lefrancais E, Roga S, Gautier V, et al. IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G. Proc Natl Acad Sci U S A 2012;109:1673-8.
20. Lefrancais E, Duval A, Mirey E, et al. Central domain of IL-33 is cleaved by mast cell proteases for potent activation of group-2 innate lymphoid cells. Proc Natl Acad Sci U S A 2014;111:15502-7.
21. Haraldsen G, Balogh J, Pollheimer J, et al. Interleukin-33 - cytokine of dual function or novel alarmin? Trends Immunol 2009;30:227-33.
22. Fang M, Li Y, Huang K, et al. IL33 Promotes Colon Cancer Cell Stemness via JNK Activation and Macrophage Recruitment. Cancer Res 2017;77:2735-2745.
23. Cayrol C, Girard JP. IL-33: an alarmin cytokine with crucial roles in innate immunity, inflammation and allergy. Curr Opin Immunol 2014;31:31-7.
24. Martin NT, Martin MU. Interleukin 33 is a guardian of barriers and a local alarmin. Nat Immunol 2016;17:122-31.
25. Lingel A, Weiss TM, Niebuhr M, et al. Structure of IL-33 and its interaction with the ST2 and IL-1RAcP receptors--insight into heterotrimeric IL-1 signaling complexes. Structure 2009;17:1398-410.
26. Roussel L, Erard M, Cayrol C, et al. Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A-H2B acidic pocket. EMBO Rep 2008;9:1006-12.
27. Ali S, Mohs A, Thomas M, et al. The dual function cytokine IL-33 interacts with the transcription factor NF-kappaB to dampen NF-kappaB-stimulated gene transcription. J Immunol 2011;187:1609-16.
28. Shao D, Perros F, Caramori G, et al. Nuclear IL-33 regulates soluble ST2 receptor and IL-6 expression in primary human arterial endothelial cells and is decreased in idiopathic pulmonary arterial hypertension. Biochem Biophys Res Commun 2014;451:8-14.
29. Choi YS, Park JA, Kim J, et al. Nuclear IL-33 is a transcriptional regulator of NF-kappaB p65 and induces endothelial cell activation. Biochem Biophys Res Commun 2012;421:305-11.
30. Sundlisaeter E, Edelmann RJ, Hol J, et al. The alarmin IL-33 is a notch target in quiescent endothelial cells. Am J Pathol 2012;181:1099-111.
31. Tsuda H, Komine M, Tominaga SI, et al. Identification of the promoter region of human IL-33 responsive to induction by IFNgamma. J Dermatol Sci 2017;85:137-140.
32. Fu J, Lin SH, Wang CJ, et al. HMGB1 regulates IL-33 expression in acute respiratory distress syndrome. Int Immunopharmacol 2016;38:267-74.

33. Islam MS, Horiguchi K, Iino S, et al. Epidermal growth factor is a critical regulator of the cytokine IL-33 in intestinal epithelial cells. Br J Pharmacol 2016;173:2532-42.
34. Richards CD, Izakelian L, Dubey A, et al. Regulation of IL-33 by Oncostatin M in Mouse Lung Epithelial Cells. Mediators Inflamm 2016;2016:9858374.
35. Polumuri SK, Jayakar GG, Shirey KA, et al. Transcriptional regulation of murine IL-33 by TLR and non-TLR agonists. J Immunol 2012;189:50-60.
36. James R, Kazenwadel J. Homeobox gene expression in the intestinal epithelium of adult mice. The Journal of biological chemistry 1991;266:3246-51.
37. Suh E, Chen L, Taylor J, et al. A homeodomain protein related to caudal regulates intestine-specific gene transcription. Mol Cell Biol 1994;14:7340-51.
38. Coskun M, Olsen AK, Holm TL, et al. TNF-α-induced down-regulation of CDX2 suppresses MEP1A expression in colitis. Biochimica et biophysica acta 2012;1822:843-51.
39. Coskun M, Soendergaard C, Joergensen S, et al. Regulation of Laminin gamma2 Expression by CDX2 in Colonic Epithelial Cells Is Impaired During Active Inflammation. J Cell Biochem 2017;118:298-307.
40. Mitchelmore C, Troelsen JT, Spodsberg N, et al. Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. The Biochemical journal 2000;346 Pt 2:529-35.
41. Boyd M, Hansen M, Jensen TGK, et al. Genome-wide analysis of CDX2 binding in intestinal epithelial cells (Caco-2). The Journal of biological chemistry 2010;285:25115-25.
42. Olsen AK, Coskun M, Bzorek M, et al. Regulation of APC and AXIN2 expression by intestinal tumor suppressor CDX2 in colon cancer cells. Carcinogenesis 2013;34:1361-9.
43. Coskun M, Olsen AK, Bzorek M, et al. Involvement of CDX2 in the cross talk between TNF- and Wnt signaling pathway in the colon cancer cell line Caco-2. Carcinogenesis 2014;35:1185-1192.
44. Troelsen JT, Mitchelmore C, Spodsberg N, et al. Regulation of lactase-phlorizin hydrolase gene expression by the caudal-related homoeodomain protein Cdx-2. Biochem J 1997;322 (Pt 3):833-8.
45. Bhat AA, Sharma A, Pope J, et al. Caudal homeobox protein Cdx-2 cooperates with Wnt pathway to regulate claudin-1 expression in colon cancer cells. PLoS One 2012;7:e37174-e37174.
46. Pinto R, Hansen L, Hintze J, et al. Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression. Nucleic acids research 2017;40:439-452.
47. Seidelin JB, Horn T, Nielsen OH. Simple and efficient method for isolation and cultivation of endoscopically obtained human colonocytes. Am J Physiol Gastrointest Liver Physiol 2003;285:G1122-8.
48. Boyd M, Bressendorff S, Møller J, et al. Mapping of HNF4alpha target genes in intestinal epithelial cells. BMC gastroenterology 2009;9:68-68.

49. Mathelier A, Fornes O, Arenillas DJ, et al. JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Research 2016;44:D110-D115.
50. Salari K, Spulak ME, Cuff J, et al. CDX2 is an amplified lineage-survival oncogene in colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America 2012;109:E3196-E3205.
51. Coskun M, Boyd M, Olsen J, et al. Control of intestinal promoter activity of the cellular migratory regulator gene ELMO3 by CDX2 and SP1. Journal of cellular biochemistry 2010;109:1118-28.
52. Verzi MP, Shin H, He HH, et al. Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Developmental cell 2010;19:713-26.
53. Verzi MP, Shin H, San Roman AK, et al. Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol Cell Biol 2013;33:281-292.
54. San Roman AK, Aronson BE, Krasinski SD, et al. Transcription Factors GATA4 and HNF4A Control Distinct Aspects of Intestinal Homeostasis in Conjunction with Transcription Factor CDX2. Journal of Biological Chemistry 2015;290:1850-1860.
55. Olsen AK, Boyd M, Danielsen ET, et al. Current and emerging approaches to define intestinal epithelium-specific transcriptional networks. American journal of physiology. Gastrointestinal and liver physiology 2012;302:G277-86.
56. Coskun M, Troelsen JT, Nielsen OH. The role of CDX2 in intestinal homeostasis and inflammation. Biochimica et biophysica acta 2011;1812:283-9.
57. Olsen J, Espersen ML, Jess P, et al. The clinical perspectives of CDX2 expression in colorectal cancer: a qualitative systematic review. Surg Oncol 2014;23:167-76.
58. Dalerba P, Sahoo D, Paik S, et al. CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. N Engl J Med 2016;374:211-22.
59. Miller AM. Role of IL-33 in inflammation and disease. Journal of inflammation (London, England) 2011;8:22-22.
60. Akimoto M, Takenaga K. Role of the IL-33/ST2L axis in colorectal cancer progression. Cell Immunol 2018.
61. Seo DH, Che X, Kwak MS, et al. Interleukin-33 regulates intestinal inflammation by modulating macrophages in inflammatory bowel disease. Sci Rep 2017;7:851.

Manuscript III

2.3 Paper III Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression

Rita Pinto[1,2,3,4], Lars Hansen[4], John Hintze[4], Raquel Almeida[1,2,3,5], **Sylvester Larsen[6,7]**, Mehmet Coskun[8,9], Johanne Davidsen[6], Cathy Mitchelmore[6], Leonor David[1,2,3], Jesper Thorvald Troelsen[6] and Eric Paul Bennett[4,*]

[1]i3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [2]Ipatimup, Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal, [3]Faculty of Medicine of the University of Porto, Porto, Portugal, [4]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark, [5]Department of Biology, Faculty of Sciences of the University of Porto, Porto, Portugal, [6]Department of Science and Environment, Roskilde University, Roskilde, Denmark, [7]Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark, [8]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark and [9]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark

Nucleic Acids Research 2017 Jul 27;45(13):e123. doi: 10.1093/nar/gkx371

The paper is a multi-site collaboration effort with several main authors. My contribution as co-author was primarily to produce, interpret, and present protein data from induction experiments performed in LS174T^{wt} and CDX2 PrIITE cells (suppl. Figure 4).

The aim of the study:

In paper III the aim was to develop a method to integrate inducible transcriptional elements into cells, in a precise manner, to enable absolute control of gene expression with no leakiness. Another aim was to demonstrate the feasibility of the method by engineering an epithelial colon cell line (LS174T) with a bi-allelic CDX2 knockout that contains a CDX2 gene with doxycycline-inducible expression. This model system was used to investigate CDX2 inducibility and leakiness and to detect novel CDX2 target genes by which expression is regulated by CDX2.

Key findings:

- The study demonstrates that the PrIITE system enables induction of gene expression while removing leakiness observed in other Tet-on systems and that it requires the use of tetracycline dosages below cell stress inducing concentrations.
- A novel colonic epithelial cell line with a bi-allelic CDX2 knockout and an inserted inducible doxycycline controlled CDX2 gene was engineered, developed, and tested.
- Novel CDX2 downstream effector genes were uncovered using the CDX2 inducible colon cancer cells generated by the PrIITE approach.
- The study provides a strategy for characterization of dose-dependent effector functions of essential genes that require an absence of endogenous gene expression.
- The established PrIITE colonic cell model system revealed that CDX2 predominantly acts as a transcriptional enhancer and rarely as a repressor.

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx371

Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression

Rita Pinto[1,2,3,4], Lars Hansen[4], John Hintze[4], Raquel Almeida[1,2,3,5], Sylvester Larsen[6,7], Mehmet Coskun[8,9], Johanne Davidsen[6], Cathy Mitchelmore[6], Leonor David[1,2,3], Jesper Thorvald Troelsen[6] and Eric Paul Bennett[4,*]

[1]i3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [2]Ipatimup, Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal, [3]Faculty of Medicine of the University of Porto, Porto, Portugal, [4]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark, [5]Department of Biology, Faculty of Sciences of the University of Porto, Porto, Portugal, [6]Department of Science and Environment, Roskilde University, Roskilde, Denmark, [7]Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark, [8]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark and [9]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark

Received October 03, 2016; Revised March 30, 2017; Editorial Decision April 10, 2017; Accepted April 27, 2017

ABSTRACT

Tetracycline-based inducible systems provide powerful methods for functional studies where gene expression can be controlled. However, the lack of tight control of the inducible system, leading to leakiness and adverse effects caused by undesirable tetracycline dosage requirements, has proven to be a limitation. Here, we report that the combined use of genome editing tools and last generation Tet-On systems can resolve these issues. Our principle is based on precise integration of inducible transcriptional elements (coined PrIITE) targeted to: (i) exons of an endogenous gene of interest (GOI) and (ii) a safe harbor locus. Using PrIITE cells harboring a GFP reporter or CDX2 transcription factor, we demonstrate discrete inducibility of gene expression with complete abrogation of leakiness. CDX2 PrIITE cells generated by this approach uncovered novel CDX2 downstream effector genes. Our results provide a strategy for characterization of dose-dependent effector functions of essential genes that require absence of endogenous gene expression.

INTRODUCTION

Historically, analysis of the molecular genetic mechanisms underlying cell fate and animal phenotypes has been studied by abrogating gene function in cellular and animal model systems. Initially this has been accomplished by random mutagenesis (1–3), homologous recombination (4) and recently by the use of precise genome editing technologies that allow for target inactivation of any GOI in cells, tissues and animal models (5–7). However, the fact that the genetic lesions induced are static render these approaches inadequate in situations where swift reversal of gene function is desired or in cases where the GOI plays an essential function for cellular survival. Thus, alternative approaches have to be employed in these situations. One commonly used alternative is based on gene 'knock down' by RNAi/shRNA (8). Although the successful application of these technologies in cell lines is well documented, knock down strategies are hampered by lack of quantitative and absolute inactivation of gene function, which makes this approach problematic in situations where downstream gene functional studies require complete gene inactivation. In these situations, the use of inducible gene expression systems has shown to be a powerful methodology that allows for: (i) control of gene expression levels of potentially toxic gene products that could have adverse side effects on cell growth and survival when expressed constitutively, (ii) temporal and spatially controlled activation of genes and proteins and (iii) analysis of cellular gene dose/response effects. Various inducible gene ex-

*To whom correspondence should be addressed. Tel: +45 35326630; Email: epb@sund.ku.dk
Present address: Rita Pinto, Department of Biosciences, Faculty of Mathematics and Natural Sciences, University of Oslo, Blindern, 0316 Oslo, Norway.

pression systems have been described in the literature based on cre-lox P system (9), myxovirus resistance 1 promoter (10), estrogen receptor (11), optogenetics (12), ecdysone-induciblity (13) or tetracyclin (Tet)-Off/On systems (14–17). The latter systems are probably the most commonly used inducible systems for which a plurality of reagents have been developed, published and are commercially available (18). However, one disadvantage of Tet-inducible systems and the majority of the aforementioned inducible systems, is their well described 'leakiness' (19–23). Furthermore, inducible expression in both transient and stable expressing cells have indicated that cells respond differently to induction, which has been attributed to heterogeneity in chromosomal integration of the inducible gene elements in individual cells leading to non-homogenous induction responses (21). Importantly, well known side effects of tetracycline (and its derivatives) on cell fitness, in particular after long term treatment, caution for its use in biomedical research when used at traditional concentrations >100 ng/ml (24–26).

Thus, there is a yet unmet need in the field to improve the tightness of the available inducible systems. We reasoned that the observed 'leakiness' with the most commonly used 'Tet-On' system (17,27) is due to the uncontrolled randomness of integration of the genetic elements encoding both the ectopically expressed transactivator and the inducible GOI, leading to muddled inducible gene expression with pleiotropic downstream effects dependent on the activation conditions used. Therefore, we hypothesize that by integrating a defined number of transactivator and inducible transcriptional elements at defined cellular genomic loci we can (i) circumvent the 'leakiness' issue and (ii) lower the Dox concentration needed for induction below the levels causing cellular stress. Thus, by taking advantage of precise genome editing and last generation Tet-On platforms (18,27), the objectives of this study were to establish a flexible 'non-leaky', minimal Dox concentration requiring isogenic knockout-rescue system. We built our cellular model system on the colorectal cell line LS174T and by mono- or bi-allelic targeting of constitutively expressing transactivator (Tet3G) elements (TET3G) to one locus and inducible GFP-reporter elements to another safe harbor locus we demonstrate minimal Dox requirement, no leakiness and reversibility of the system. We next demonstrate the utility of the isogenic PrIITE system by mono- or bi-allelic integration of inducible CDX2 transcription factor elements into PrIITE cells and re-confirm the reversibility and lack of leakiness of the system. We also for the first time demonstrate, that Tet-On leakiness is related to the cellular copy number of integrated transactivator elements and not the number of integrated inducible elements of the inducible system.

Finally we confirm the complete absence in leakiness of the system by RNA-seq and in combination with ChIP-seq identify several novel genes directly transcriptionally controlled by CDX2 (Figure 1). The selection of CDX2 as our target gene was based on its relevance as a key regulator of intestinal differentiation, with many downstream targets that can be assessed as read-outs (28).

MATERIALS AND METHODS

ZFN gene targeting plasmids and plasmid donor construction

CDX2 and AAVS1 CompoZr ZFN plasmids for human *CDX2* and *AAVS1* were obtained from Sigma targeting exon 1 of *CDX2* (AACTTCGTCAGCCCCccgcagTACCC GGACTACGGCGGTT) and intron one of the PPP1R12C gene at the *AAVS1* hotspot AAV integration site (ACCCCA CAGTGGggccacTAGGGACAGGAT), respectively. ZFN binding sites are shown in upper case and linker cut site in lower case lettering. ZFNs were tagged with 2A peptide fused GFP or Crimson as described recently (29). Fluorescent protein tagging allows for FACS enrichment and improves efficiency in obtaining correctly targeted clones.

Donor construct was designed based on the previously described approach by Maresca *et al.* (30) with the only modification that inverted ZFN binding sites were positioned flanking the entire donor insert. A synthetic ObLiGaRe CDX2 donor vector frame work possessing inverted CDX2 ZFN binding sites (AACCGCCG TAGTCCGGGTAccgcagGGGGCTGACGAAGTT) flanking a XhoI/EcoRV linker was generated (EPB64, Genewiz, USA, Addgene ID#90017). A CMV-Tet3G-SV40UTR TET3G transactivator encoding fragment (pTet3G, Clontech/Takara, USA) was XhoI/HindIII excised (HindIII overhang bluntended) and inserted into XhoI/XbaI (overhangs bluntended) site of EPB64-donor vector generating *pCDX2-pCMV-TET3G-ObLiGaRe* donor vector.

Various inducible gene expression ObLiGaRe donor constructs targeted to the *AAVS1* safe harbor locus were generated based on available *AAVS1* safe harbor CompZr ZFN binding element information (Sigma-Aldrich, USA). A *AAVS1* ObLiGaRe donor vector framework was generated (EPB58, Genewiz, USA, Addgene ID#90016) and designated pObLiGaRe-*AAVS1*, possessing the respective inverted ZFN binding elements (5′-atcctgtcccta ggccacccactgtgggt-3′) flanking an EcoRV multiple cloning site linker (see Supplementary Figure S11 for vector design). A full-length codon optimized CDX2 (CDX2opt) construct was generated (EPB40, GeneArt/ThermoFisher, USA and SalI/BamHI insert cloned into SalI/BamHI sites of pTRE3G (Clontech/Takara, USA) generating pTRE3G-CDX2opt.

A BamHI fragment encoding a Golgi targeted EYFP fusion protein, T2EYFP (31) was blunt end cloned into the EcoRV site of pTRE3G generating *p*TRE3G-T2EYFP. T2EYFP encodes the N-terminal Golgi targeting and retention sequence fused to EYFP. The XbaI/XhoI Tet responsive expression fragments from both *p*TRE3G-CDX2opt and *p*TRE3G-T2EYFP were excised and blunt end cloned into the EcoRV site of pObLiGaRe-AAVS1 (EPB58) generating *p*AAVS1-TRE3G-CDX2opt and *p*AAVS1-TRE3G-T2EYFP. For *p*AAVS1-TRE3G-T2EYFP, the donor insert was excised with PacI/PmeI and inserted into PacI/PmeI site of a modified AAVS1 targeting vector (EPB71, Genewiz, USA, Addgene ID#90018) where PacI/PmeI cloning sites are flanked by insulator sequences and the inverted AAVS1 ZFN binding sites. See Supplementary Fig-

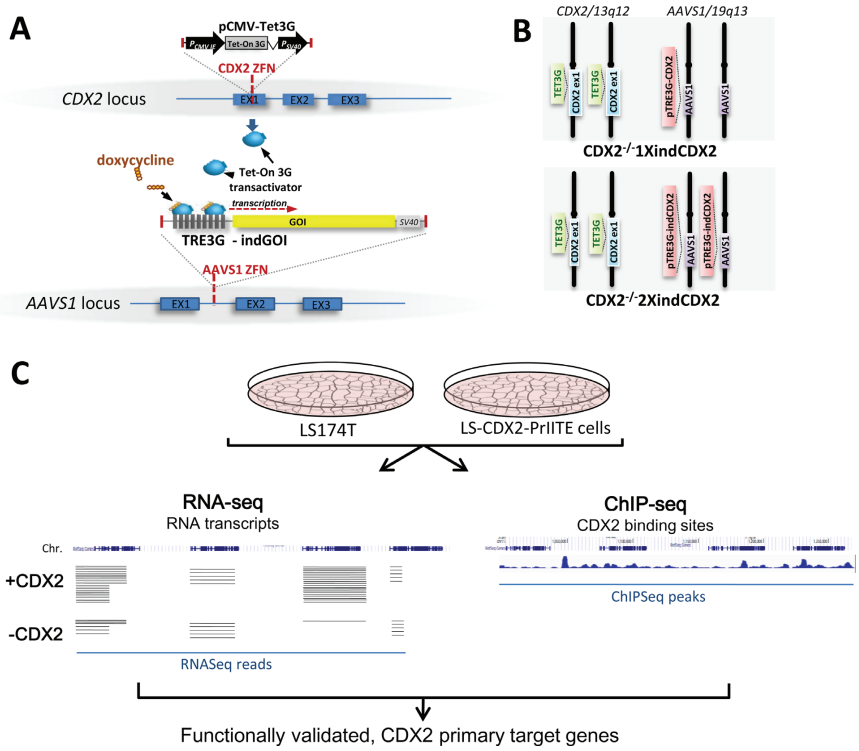


Figure 1. Overview of the established isogenic knockout-rescue model system based on precise integration of inducible elements (PrIITE) in LS174T. (**A**) pCDX2-pCMV-TET3G-ObLiGaRe donor vector (pCMV-Tet3G) was targeted to CDX2 exon1 by ZFN mediated targeted integration followed by pTRE3G-inducible Gene Of Interest (GOI) ObLiGaRe donor vector (pTRE3G-indGOI) targeting to the safe harbor AAVS1 locus. (**B**) Schematic illustration of the isogenic inducible CDX2 LS174T cell model system based on the mono- (upper panel) or bi- (lower panel) allelic precise genome targeting of CDX2 inducible elements, CDX2$^{-/-}$1XindCDX2 and CDX2$^{-/-}$2XindCDX2 respectively. (**C**) Illustration of the experimental workflow including deep transcriptomic analysis (RNA-seq) of wild type LS174T and isogenic indCDX2-PrIITE cells exposed to variable concentrations of doxycycline inducer and CDX2 target validation by genome-wide CDX2 ChIP-seq (ChIP-seq) analysis.

ure S11 for vector constructs used and generated. All plasmids were Sanger sequencing verified.

Cell culture and transfection

LS174T (ATCC-CL-188) human intestinal cell line was maintained in Ham's nutrient mixture F12/Dulbecco's modified Eagle's medium (1:1) supplemented with 10% fetal bovine serum and 1% glutamine. Different target 'knock in' strategies were undertaken, including use of 800 bp homology arms flanking the donor integration cassette D (data not shown). Successful LS174T targeted KI was only obtained when the ObLiGaRe strategy for improved target specific 'knock in' integration of donor constructs was employed (30). In brief, the ObLiGaRe strategy is based on use of existing ZFN binding elements that in an 'inverted' orientation flank the donor DNA construct. By co-nucleofection of ZFN's and donor DNA, double stranded breaks at both the specific chromosomal target site and flanking the donor plasmid will occur allowing cellular repair pathway mediated target specific integration of the linearized donor at the desired target site.

For stable *pCDX2-pCMV-TET3G-ObLiGaRe* integration the first exon of the *CDX2* gene was targeted through CDX2 CompoZr ZFNs driven integration (Sigma-Aldrich, USA) (AACTTCGTCAGCCCCccgcagTACCCGGACTACGGCGGTT, left and right ZFN binding sites capitalized and linker cloning site in lower case). 1×10^6 LS174T cells were transfected by nucleofection, simultaneously with the GFP/E2 Crimson tagged CDX2 ZFN plasmids (2 μg each) and the pCDX2-CMV-TET3G-donor vector (5 μg). Nucleofector solution T (Lonza, CH)/Nucleofector program T-020 were used for the electroporation procedure in an Amaxa Cell Line Nucleofector device (Lonza, CH). Cells were then cultured for 6 h at 37°C to stabilize, and then moved on to a 30°C cold shock. Two days after transfection, cells expressing both GFP and E2 Crimson were sorted out in a FACS ARIA III (BD BioSciences, USA) as previously described (29). The cell bulk was plated out to grow in collagen-coated plates and two rounds of cloning were then employed.

Single cell clones were screened by junction PCR using a primer flanking the 5′ *CDX2* target locus and a reverse oriented primer localized within the integrated *pCMV-TET3G* cassette (Supplementary Figure S1A and B). Based on IDAA assay (32), one of them (5B5) showed the presence of wt allele, representing heterozygozity for *p*CMV-TET3G integration at the *CDX2* locus (clone named $LS^{CDX2+/-}$), while the other two clones (3D6 and 5E2) were found to possess successfully integrated *p*CMV-TET3G transactivator at both alleles representing homozygozity for *p*CMV-TET3G (clones named $LS^{CDX2-/-}$#1 and $LS^{CDX2-/-}$#2 respectively).

1×10^6 cells from a *CDX2* KO clone containing mono or bi-allelic *p*CMV-TET3G transactivator inserted into the two *CDX2* alleles were then transfected by nucleofection with GFP/E2 Crimson tagged AAVS1 ZFN plasmids (2 μg each) and 5 μg *p*AAVS1-TRE3G-CDX2opt or *p*AAVS1-TRE3G-T2EYFP.

Clone characterization by polymerase chain reaction (PCR) and IDAA

In order to identify correct integration of *pCDX2*-CMV-TET3G, a region comprising a part of *CDX2* exon 1 upstream the ZFNs cutting site and left 800 bp homology arm and a part of the CMV promoter was amplified using Expand Long Template PCR System (Roche Applied Science, GE) (5′-CMV-TET3G junction PCR). Similarly, the identification of correctly integrated *p*AAVS1-TRE3G-CDX2opt or *p*AAVS1-TRE3G-T2EYFP fragments into *AAVS1* locus was performed by amplification of a region comprising a part of *AAVS1* upstream the ZFNs cutting site and a part of the *TRE* promoter (5′-TRE3G-CDX2opt or TRE3G-T2EYFP junction PCR). PCR reaction mixture consisted of 100 ng DNA, 2.5 μl buffer 1, 3.0 μl 1.25 mM dNTPs mix, 0.25 μl each primer at a concentration of 25 μM, 0.1 μl enzyme mix and water to a final volume of 25 μl. All junction PCR primers were obtained from TAG Copenhagen A/S, Denmark and are listed in Supplementary Table SIV. Amplification was done using the following touch down protocol. After preheating for 5 min at 95°C, 12 cycles were performed starting with denaturation for 45 s at 95°C, annealing for 15 s at 74°C with a decrease in annealing temperature of –1°C/cycle, and 2 min at 72°C, followed by an additional 25 cycles of 45 s at 95°C, 15 s at 64°C and 2 min at 72°C, followed by a final extension of 3 min at 72°C. PCR products were run in a 1.2% agarose gel, bands were gel purified and sequence confirmed by Sanger sequencing.

FACS-enriched stable clones (29) were screened by junction PCR using primers flanking the junction between the *CDX2* or *AAVS1* genes and the integrated cassette (Supplementary Figures S1 and S2).

Presence of unmodified CDX2 or AAVS1 target (WT allele presence test) was performed using the recently described IDAA method (32) and protocol guidelines (33). In brief, due to size of the donor constructs used (>2 kb, respectively) the respective CDX2 or AAVS1 ZFN target sites can only be successfully amplified if integration has not occurred at the respective target loci. PCR was performed using 100 ng DNA in 25 μl using AmpliTaq Gold (ABI/Life Technologies) for *CDX2* locus or TEMPase Hot Start DNA Polymerase (Ampliqon A/S, DK) for *AAVS1* locus using the recently described IDAA/tri-primer amplification conditions comprising primers flanking the respective target site in combination with a universal 6-FAM 5′-labeled primer (FamF), specific for a 5′-overhang attached to the forward primer. IDAA primers used were purchased from TAG Copenhagen A/S and are listed in Supplementary Table SIV. Fluorescently labeled amplicons were then analyzed by capillary electrophoresis based fragment analysis using an ABI3030 instrument (Applied Biosystems/Life Technologies, USA). Raw data obtained was analyzed using Peak Scanner Software V1.0 (Applied Biosystems/Life Technologies, USA).

Induction with Dox

The Tet-On3G system was induced with Dox (Sigma-Aldrich, USA) in a range between 0.004 and 4 μg/ml for CDX2 knock-in (KI) clones and 0.001–5 μg/ml for T2EYFP KI clones. Dox was added to the medium every 24 h for one or 2 days, and medium was exchanged every 48 h after Dox removal. Controls where Dox was not added to the medium were used (0 μg/ml of Dox).

Western blot of cell lysates

Whole-cell extracts were obtained by resuspension of cell pellets in RIPA buffer (50 mM Tris–HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.1% sodium dodecyl sulphate) in the presence of complete protease inhibitors cocktail (Roche Applied Science, GE). Quantification of total protein was determined by bicinchoninic acid protein assay (Thermo Scientific™ Pierce, USA). 15 μg of protein extracts were then analyzed by standard SDS-PAGE, transferred to a nitrocellulose membrane (GE Healthcare Life Sciences, USA) and blotted on at 4°C with mouse monoclonal primary antibodies CDX2-88 1:500 (Biogenex, USA), anti-TetR 1:1000 (Clontech/Takara, USA), anti-MUC2, undiluted hybridoma supernatant and anti-actin 1:8000 (Santa Cruz Biotechnology, USA) in 5% BSA in PBS. Membranes designed for MUC2 blotting were previously treated with neuraminidase from *Clostridium perfringens* type VI (Sigma-Aldrich, USA) diluted in PBS to a final concentration of 0.2 U/ml, for 1 h at 37°C. Peroxidase-conjugated secondary antibodies (goat polyclonal anti-mouse-HRP 1:2000 for CDX2, TetR and MUC2 and goat polyclonal anti-rabbit-HRP 1:2000 for actin (Santa Cruz Biotechnology, USA) were used and developed with the ECL detection kit (Bio-Rad Laboratories, USA).

RNA extraction

Total RNA was extracted using RNeasy kit (Qiagen, GE) as recommended by the manufacturer.

Transcriptomics (RNA-Seq)

Polyadenylated RNA was isolated from total RNA using standard protocols (Dynabeads mRNA Direct Micro Kit, Ambion/Life Technologies, USA) starting with 5 μg total RNA. Library preparation for NGS using Ion Torrent technology (Life Technologies, USA) was carried out according

to the manufacturers recommendations (Ion Total RNA-Seq Kit v2, Ion torrent/Life Technologies, USA), and sequenced using an Ion Proton system (Life Technologies, USA). Quality control, quantification of RNA and libraries was carried out using using Agilent RNA 6000 Nano Kit or Agilent High Sensitivity DNA kit and Agilent Bioanalyzer (Agilent Technologies, USA). Sequencing reads were mapped to hg19 and bioinformatics analysis was conducted using CLCs Genomic Workbench (CLC bio/Qiagen, DK), STARBowtie2 followed by Cufflinks for expression analyses. In brief, ∼50 ng polyA RNA was fragmented down to 100–300 base fragments using RNaseIII for 1–3 min followed by adapter ligation, amplification for 9–14 cycles and barcoding using Ion Express RNA-Seq Barcode kit (Ion Torrent/Life Technologies, USA). The final library fragment size and concentration was determined by Agilent Bioanalyzer analysis followed by template preparation using Ion PI Template OT2 200 Kit v3 (Life Technologies, USA) and Ion One Touch System followed by NGS on an Ion Proton system using Ion PI™ Chip Kit v2. In general, two transcriptome libraries were barcoded and analyzed on one Ion PI v2 chip. Sequencing depth for the RNA-Seq data sets ranged from 34.5 to 44.6 mill reads/sample with 95.4–97.7% mappable reads to the hg19 reference data base. Mean read length for transcriptome RNA-Seq data ranged from 116 to 136 bp.

Genome-wide analysis of CDX2-binding sites by ChIP-seq.

LS174T cells grown for five days in a 30 × 30 cm culture dish were cross-linking and sonicated as described previously to generate fragments of ∼0.2 to 1.2 kb. Hereafter, the ChIP-protocol was performed as previously reported (34). Briefly, immunoprecipitation was done in four replicates and performed overnight at 4°C using specific antibodies to human CDX2 (CDX2-88, Biogenex, USA) and the influenza hemagglutinin (HA) epitope (rabbit polyclonal α-HA; Santa Cruz Biotechnology Inc, Heidelberg, Germany), used as a negative control. Immunocomplexes were recovered with 50 μl protein A/G beads (Invitrogen/ThermoFisher Scientific, USA). Verification of the enrichment in the CDX2-immunoprecipitated DNA samples was done by qPCR analyzing the DNA level from known CDX2-targets in the CDX2-immunoprecipitated samples and comparison it to level in the negative control (HA-immunoprecipitated DNA).

In order to increase the depth of the analysis, the ChIP library established was deep sequenced using Ion Proton instrumentation generating an approximate ≈54 million reads depth. Library preparation for NGS was carried out with the Ion Xpress Plus gDNA Fragment Library Kit according to User Bulletin 4473623: Ion Chipseq Library Preparation on the Ion Proton System (Life Technologies, USA), starting with 10 ng DNA. Quality control of the libraries was carried out using using Experion DNA 1K analysis kit and the Experion System (Bio-Rad Laboratories, USA). In brief, 10 ng DNA was end-repaired and ligated to barcode adaptors from the Ion Xpress Barcode Adapters 1–16 Kit (Life Technologies, USA), followed by nick repair and amplification for 18 cycles. The amplified libraries were subjected to two rounds of bead capture with the Agencourt AMPure XP Kit (Beckman Coulter, USA) to size-select fragments ∼160–340 bp in length. The final library concentration was determined by Qubit analysis (ThermoFisher Scientific, USA). The barcoded ChIP- and IP control-DNA-libraries were handled similarly to the RNA-Seq library and analyzed on a single Ion PI™ Chip v2. For the ChIP- and IP control-DNA-libraries a total of 54.5 mill reads (21.8 and 32.7 mill respectively) were obtained and 98.5% were mappable to the reference hg19 data base with a mean read length of 130 bp. The detection of CDX2 ChIP seq peaks and the location of the closest gene was performed using CisGenome version 2 (35).

Immunodetection by fluorescence-activated cell-scanning (FACS) analysis

Cells were trypsinized at different time-points after induction with Dox or after Dox removal and washed twice with PBS. After being centrifuged at 1200 rpm for 7 min, CDX2 KI cells were fixed in 4% PFA for 20 min., washed in PBS and permeabilized in cold methanol for 15 min. After washing with PBS, samples were incubated with the primary antibody (CDX2-88 clone, BioGenex, 1:500.) for 1 h at 37°C. Cells were then stained with FITC-conjugated AfinityPure immunoglobulin antimouse IgG (Jackson ImmunoResearch Laboratories, USA) diluted 1:100 in 0.05% BSA in PBS and then subjected to FACS in a FACS ARIA III (BD BioSciences, USA). T2EYFP KI cells were immediately scanned by FACS after cell trypsinization.

Immunocytochemistry

Cell slides were fixed in cold acetone for 15 min. For TetR anti-TetR cells were incubated overnight at 4°C with the primary antibodies (9G9 clone, Clontech/Takara, USA). Negative controls were performed by omission of primary antibodies. After washing, a rabbit anti-mouse Ig FITC-labeled secondary antibody (Dako A/S, DK) diluted 1:100 in PBS with 5% BSA was added for 45 min, protected from light. DAPI was used as a nuclear counterstain and slides were mounted in Vectashield mounting media (Vector Labs, USA).

For EYFP visualization on the induced T2EYFP KI clones, trypsinized cells were fixed in cold acetone for 10 min, DAPI was added and slides mounted. Samples were examined under a Zeiss fluorescence microscope equipped with DAPI and FITC interference filters. Images were acquired using a Zeiss Axioskop 2 and an AxioCam MR3 camera and Zeiss Application Suit software.

Real time *in vivo* video material

For the time lapse video, 1×10^6 CDX2$^{-/-}$2XindT2EYFP cells were seeded in a well of a six-well plate and 0.5 μg/ml of Dox were added after 24 h. Time lapse images started to be acquired under dark field immediately after Dox addition and every 20 min during 24 h using a Leica DMI 6000 timelapse microscope equipped with FITC interference filter and with CO_2 supply.

RESULTS

Precise integration of inducible transcriptional elements (PrIITE)

CDX2 knock-out by targeted integration of Tet3G transactivator elements. In order to eliminate the adverse cellular effects seen from prolonged exposure to Dox (25,26), we designed a strategy combining stable and controlled TeT-On transactivator/TeT3G expression with disruption of a target gene. First we established a Tet3G expressing LS174T cell line, in which the TET3G elements were targeted to *CDX2* exon1 CDS, thereby abrogating the endogenous gene function. Targeting *CDX2*, a master differentiation transcription factor (36), allowed us to determine the efficacy of the system in an unprecedented CDX2 knock out cell system (Figure 1). Multiple attempts at homologous recombination driven donor integration at the CDX2 exon 1 locus based on CDX2 ZFNs and a TET3G donor template flanked by 700bp homology arms did not give rise to any correctly targeted clones, likely due to low homologous repair capacity of LS174T cells used in this study (data not shown). *pCDX2-pCMV-TET3G-ObLiGaRe* donor vector (Figure 1A and Supplementary Figure S1A) was co-transfected together with *CDX2* CompoZr ZFN plasmids into LS174T cells, whereby *pCMV-TET3G transactivator* donor vector was specifically integrated into exon1 of *CDX2*, as illustrated in Figure 1A and Supplementary Figure S1A and B, by ObLiGaRe driven recombination (see Material and Methods section for details). Three clones (5E2, 3D6 and 5B5) were shown to include correct integration of the *p*CMV-TET3G construct and constitutively express the Tet3G transactivator (Supplementary Figure S1C), one clone, 5B5, was mono allelically targeted (designated $LS^{CDX2+/-}$) and CDX2 protein expression maintained at levels similar to the parental LS174T wt cell line and in the bi-allelic targeted *CDX2* knock-out (KO) clones, 3D6 ($LS^{CDX2-/-}$#1) and, 5E2 ($LS^{CDX2-/-}$#2), *CDX2* expression was completely abrogated (Supplementary Figure S1C and D). Constitutive Tet3G expression and abrogation of CDX2 was maintained after prolonged cell passaging and after freeze/thawing lack of CDX2 was correlated with significantly decreased protein levels of a well-know CDX2 target gene, *MUC2* (37) (Supplementary Figure S1C and D). Lastly, we tested the functionality of the mono and bi–allelically targeted *CDX2* KO cells by cellular transfection with known CDX2 enhancer reporter constructs and could show dose dependent decreased expression of a *HNF4A* reporter and >10-fold reduced expression of a *HEPH1* reporter (38) in the *CDX2* KO cells (Supplementary Figure S3).

We thus show, that *CDX2* KO by ZFN-mediated site specific *p*CMV-TET3G integration in LS174T cells was effectively accomplished and that stable Tet3G expression was maintained over time.

Safe harbor targeted integration of inducible T2EYFP elements. Having shown that both $LS^{CDX2-/-}$#1 and #2 display similar transactivator expression levels, an inducible T2EYFP model system was established in $LS^{CDX2-/-}$#1. The inducible *p*TRE3G-T2EYFP-*ObLiGaRe* donor vector was targeted to the safe harbor *AAVS1* locus by co-transfection with *AAVS1* CompoZr ZFN's into the $LS^{CDX2-/-}$#1 cells as illustrated in Figure 2A. Multiple correctly targeted indT2EYFP clones were obtained and two of these selected for further detailed analysis, one mono-allelic T2EYFP targeted clone 7E9 (hence forward referred to as $CDX2^{-/-}$1XindT2EYFP) and another bi-allelic targeted clone 5G8 (henceforward referred to as $CDX2^{-/-}$2XindT2EYFP), Supplementary Figure S2. Both clones maintained constitutive Tet3G expression and absence of CDX2 (Figure 2B).

LS174T indT2EYFP PrIITE cells display no leakiness and Dox induction is fast and reversible. We first aimed to determine the leakiness of the inducible T2EYFP reporter in the $CDX2^{-/-}$1xindT2EYFP and $CDX2^{-/-}$2xindT2EYFP clones by FACS and immunofluorescence, which showed undetectable fluorescence in uninduced PrIITE cells (Figure 2C, D and E). Upon induction, immunofluorescence displayed the expected peri-nuclear reactivity, consistent with the correct sub-cellular localization of T2EYFP fusion protein in the Golgi apparatus (39) (Figure 2E).

We next determined the induction dynamics over time by *in vivo* live imaging, demonstrating that the induction dynamics is very rapid. T2EYFP expression is detectable within 4 h post induction and all cells display full expression of the reporter after 8–10 h (Supplementary video material). Importantly, the induction dynamics is fully synchronized and all cells simultaneously become positive for the T2EYFP reporter. To elucidate how Dox dosage might influence on the dynamics of the system, induction was performed using variable Dox concentrations ranging from 0 to 5 μg/ml. As shown in Figure 2C, modest induction was observed with dox concentrations <5 ng/ml, whereas significant induction required Dox concentration >10 ng/ml. Interestingly, these results suggest, that the induction potential is related to the number of inducible elements integrated into the system. For instance, using a 10 ng/ml Dox dosage rendered >90% of the $CDX2^{-/-}$2XindT2EYFP cells EYFP reporter positive, whereas <50% $CDX2^{-/-}$1xindT2EYFP cells were found positive for this dosage.

Next, the reversibility of the PrIITE cell system was quantified by FACS analysis of the EYFP expressing cell population after Dox removal from the media (Figure 2D). As observed, induction was fully reversible and after 10 days post Dox removal, all cells in the population returned to non-expressing levels. However, the dynamics in reversibility was found to be dependent on both number of inducible elements integrated and the Dox dosage used. Clearly reversal to non-expressing levels was achieved rapidly and within 5 days for $CDX2^{-/-}$1XindT2EYFP and slower for $CDX2^{-/-}$2xindT2EYFP using the lower 50 ng/ml induction dosage.

We thus demonstrate that the PrIITE cell system displays no leakiness, is fully reversible, and that the induction dynamics occurs fast and in a synchronized manner and require below cellular stress inducing Dox concentrations.

LS174T indCDX2 PrIITE cells show no leakiness and ectopic CDX2 expression is tightly controlled requiring minimal Dox induction levels. The role of CDX2 in intesti-

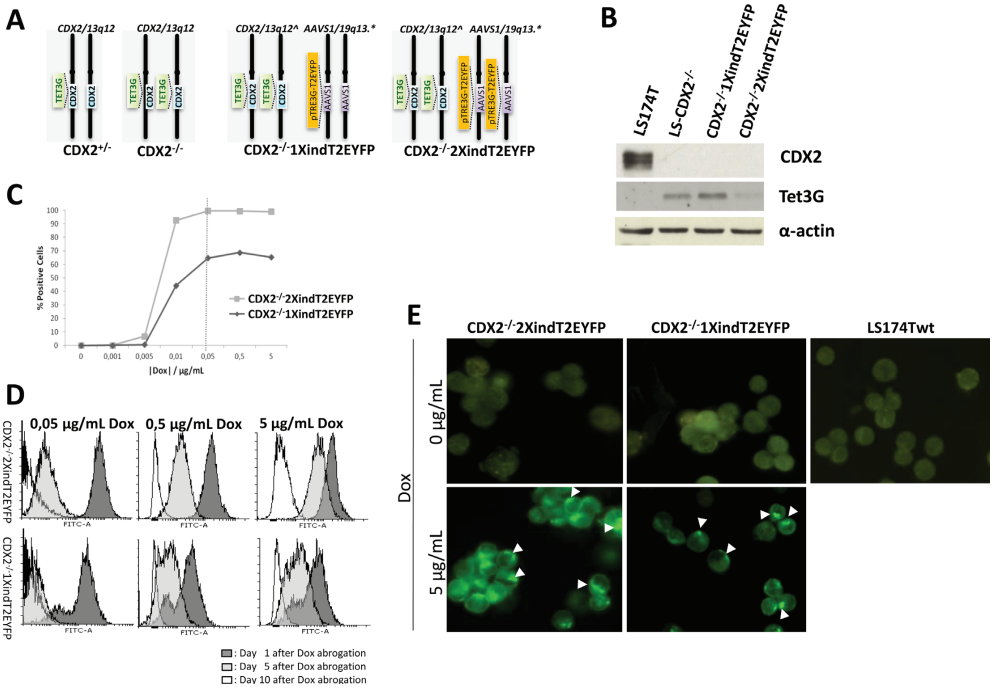


Figure 2. Overview of the proof of principle system using inducible T2EYFP reporter in isogenic LS174T-PrIITE cells. (**A**) First pCMV-Tet3G was targeted to *CDX2* exon1 of LS174T cells generating mono- or bi-allelic CDX2 KO cells, CDX2$^{+/-}$ and CDX2$^{-/-}$ respectively. Next, cells were re-targeted with *pTRE3G-T2EYFP ObLiGaRe-donor vector* (pTRE3G-T2EYFP) directed to the safe harbor *AAVS1* locus generating CDX2$^{-/-}$1XindT2EYFP and CDX2$^{-/-}$2XindT2EYFP cells. (**B**) Western blot analysis of LS174wt, CDX2$^{-/-}$, CDX2$^{-/-}$1XindT2EYFP and CDX2$^{-/-}$2XindT2EYFP cells show absence of CDX2 and stable TeT3G expression. (**C**) FACS based determination of percentage indT2EYFP positive cells after induction with different Dox concentrations. The graph shows that induction is dose and copy number dependent in CDX2$^{-/-}$1XindT2EYFP and CDX2$^{-/-}$2XindT2EYFP PrIITE cells. Approximate 100% T2EYFP positive cells was achieved using >0.005 μg/ml Dox concentration for CDX2$^{-/-}$2XindT2EYFP. Maximum 70% positivity was achieved regardless of the higher Dox induction concentrations used for CDX2$^{-/-}$1XindT2EYFP cells. Importantly, CDX2$^{-/-}$2XindT2EYFP cells show complete absence in T2EYFP fluorescence in uninduced cells thus, no leakiness of T2EYFP LS-PrIITE cells was detectable. Critical dox concentration above which adverse cellular stress (24–26) is caused is indicated by a dotted gray line. (**D**) Reversibility of induction as determined by FACS analysis of fixed cells 1, 5 or 10 days post 48 h induction with variable Dox induction concentrations. Notably, reversibility was copy number and dox dependent and full reversibility within 10 days was only achieved for CDX2$^{-/-}$1XindT2EYFP PrIITE cells using the lower 0.05 μg/ml Dox concentration. (**E**) Sub cellular localization of inducible T2EYFP as determined by fluorescence of fixed trypsinized indCDX2 PrIITE cells with or without dox induction. Arrow heads indicate expected Golgi localization of T2EYFP.

nal gene regulation is well known. However, to the best of our knowledge no studies so far described the effect of complete ablation of *CDX2* from the cellular genome. In a single study *CDX2* has been suggested as a lineage survival oncogene when amplified in colorectal cancer (40). In order to distinguish endogenous from ectopically expressed *CDX2* transcripts the inducible *CDX2* open reading frame was codon optimized (hereafter referred to as indCDX2). *pTRE3G-indCDX2-ObLiGaRe* donor vector was targeted to the *AAVS1* safe harbor site of LS$^{CDX2-/-}$#1 cells generating clones 7D9 (CDX2$^{-/-}$1XindCDX2) and 6D6 (CDX2$^{-/-}$2XindCDX2) possessing one or two inducible codon optimized CDX2 copies respectively (Figure 1B and Supplementary Figure S4). Constitutive Tet3G expression was maintained in targeted cells and CDX2$^{-/-}$2XindCDX2 or CDX2$^{-/-}$1XindCDX2 cell induction dynamics over 24 h were similar (Figure 3A and B). To test the reversibility of the CDX2 PrIITE system built, we induced with a range of Dox concentrations (0–4 μg/ml) for 48 h followed by removal of Dox from the medium. As expected, no expression of indCDX2 was revealed in non-induced cells followed by fast dose dependent Dox induction dynamics (Figure 3C). Importantly, for the lowest concentration of Dox used (0.004 μg/ml), the induction dynamics for the CDX2$^{-/-}$ 1XindCDX2 clone was slower than for CDX2$^{-/-}$2XindCDX2 (Figure 3C). Of notice, the indCDX2 protein levels, reached for CDX2$^{-/-}$1XindCDX2 with 4 ng/ml Dox dose, were similar to the endogenous CDX2 levels detected in LS174Twt cells (Supplementary Figure S4E) while for CDX2$^{-/-}$2XindCDX2 induction reached 120× higher levels relative to LS174Twt levels (Supplementary Figure S4E). In contrast to the dynamics in reversibility for indT2EYFP KI clones (Figure 2), reversibility of indCDX2 induction appeared faster and required very modest Dox induction dosages in the sub cell stress inducing nanogram/ml range.

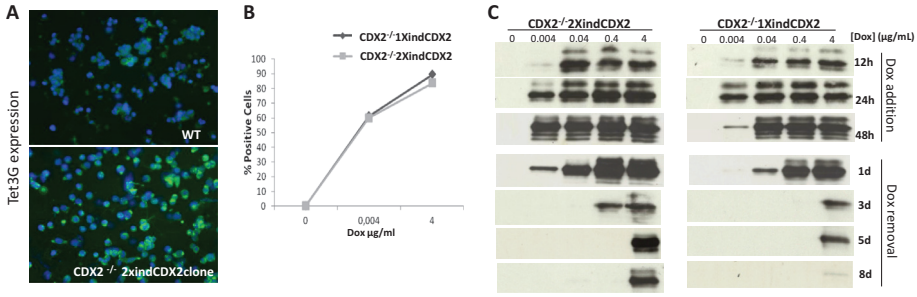


Figure 3. Induction dynamics of isogenic inducible CDX2 LS174T PrIITE cells. (**A**) Anti-TeT3G transactivator immunofluorescence in CDX2$^{-/-}$2XindCDX2 cells and not in LS174T wt cells. (**B**) 24 h induction kinetics of indCDX2 using different Dox concentrations. >60% of CDX2$^{-/-}$1XindCDX2 and CDX2$^{-/-}$2Xind CDX2 cells become indCDX2 positive after 24 h 0.004 μg/ml Dox induction. Use of higher Dox concentrations higher percentage of cells become CDX2 positive after 24 h Dox treatment. (**C**) Dynamics in reversibility of the PrIITE system using different Dox induction concentrations. CDX2$^{-/-}$2Xind CDX2 (left panel) and CDX2$^{-/-}$1Xind CDX2 (right panel) cells were induced 48 h with the Dox concentrations indicated above the panels. Hereafter Dox was removed and post induction indCDX2 protein levels were determined by Western blot analysis 1, 3, 5 or 8 days post Dox removal. Maximum indCDX2 induction levels were achieved using 0.004 μg/ml Dox concentration for CDX2$^{-/-}$2XindCDX2 whereas 0.040 μg/ml Dox was required for CDX2$^{-/-}$1Xind CDX2.

Based on western blot analysis indCDX2 protein was undetectable within 3 days post Dox removal for both mono- and bi-allelic clones (Figure 3C).

Validating the tightness in inducibility of the PrIITE by deep transcriptome analysis. To confirm the tightness of the PrIITE system and in order to determine the CDX2 downstream target genes, LS174Twt, uninduced and induced CDX2$^{-/-}$2XindCDX2 or CDX2$^{-/-}$1XindCDX2 cells were RNA-seq transcriptome profiled (Figure 4A, and Supplementary Figure S5). To ensure that the isogenic PrIITE cells generated were representative of the mosaic moderately well differentiated LS174Twt cell population (41), all comparative analysis here and in the following sections included LS174Twt cells. Approximately 10 000 genes with RPKM >1 were shown to be expressed in the wt cells and the expression profiles for the mucin genes *MUC2* and *MUC5AC* correspond to the profiles previously reported for LS174Twt cells with high expression for the former and low expression for the latter (42,43) (Figure 5B). CDX2 target genes were defined by the following criteria; genes with RPKM >1 in wt cells where RPKM values in uninduced PrIITE cells (*CDX2*$^{-/-}$) were reduced >4× (4×↓) and increased >2× (2×↑) after re-induction relatively to the uninduced RPKM values. In depth analysis of the RNASeq reads that mapped to the human exon1 *CDX2* gene locus, including the region across the *CDX2* ZFNs cutting site, confirmed the presence of out of frame transcripts possessing deletions in CDX2$^{-/-}$2XindCDX2 exon1, while the heterozygous CDX2$^{-/-}$1XindCDX2 clone expressed intact endogenous *CDX2* transcripts present at ≈40% of the LS174Twt transcript levels (Supplementary Figure S5A). Analysis of the unmapped RNA-seq reads confirmed the presence of the Tet3G transactivator transcripts in CDX2$^{-/-}$ cells (Supplementary Figure S5B) and importantly, no indCDX2 transcript was detected in the uninduced state of the CDX2$^{-/-}$2XindCDX2 clone (Supplementary Figure S5B), thus confirming the biochemical and immunofluorescence results obtained previously.

Taken together, by biochemical analysis, immunofluorescence and deep transcriptome analysis, we have shown that the PrIITE system allows for absolute control of inducible and reversible gene expression of CDX2 and it's downstream target genes.

Validating the transcriptome identified CDX2 target genes by genome-wide ChIP-seq analysis. Reasoning that the RNA-seq identified CDX2 target genes could be direct or indirect targets of CDX2, we wanted to validate the RNA-seq data set by genome wide chromatin immunoprecipitation sequence analysis (ChIP-seq). We thus used ChIP-seq to identify true CDX2 binding elements in the vicinity of the RNA-seq identified CDX2 target genes. ChIP was performed on LS174Twt cells using a well-known anti-CDX2 monoclonal antibody essentially using a previously reported procedure (38). A LS174T non-immunoprecipitated library was deep sequenced in parallel and used as ChIP-seq background control. An additional requirement was added to the RNA-seq criteria for positive scoring of primary CDX2 target genes, in that a ChIP-seq peak (CDX2 binding element) were to be present within or 1 kb up or downstream of the target gene. The combined RNA- and ChIP-seq criterion identified 31 direct CDX2 target genes (Figure 4B and Supplementary Table SI). Among these genes several known CDX2 regulated targets were found, such as *TFF3*, *MUC2* and *CDX2* itself, but several novel targets were also found, including *MUC5B, MUC5AC*, *MUC6*, *GPA33* and *LDLR* (Figure 5). Furthermore, *in vitro* promotor analysis confirmed the CDX2 dependent regulation of the enhancer identified in the novel CDX2 target gene *GPA33* (44) (Figure 6) and thus, substantiated both our RNA- and ChIP-seq findings. Interestingly, a significant intergenic CDX2 binding element was identified in the 300 Mb 11p15 region containing the mucin genes *MUC6*, *MUC2*, *MUC5B* and *MUC5AC* (Supplementary Figure S6A). Clearly our transcriptome results reveal a significant concerted down-regulation of this gene cluster and suggest CDX2 as a 'locus control gene' (Figure 5B). As expected, the non CDX2

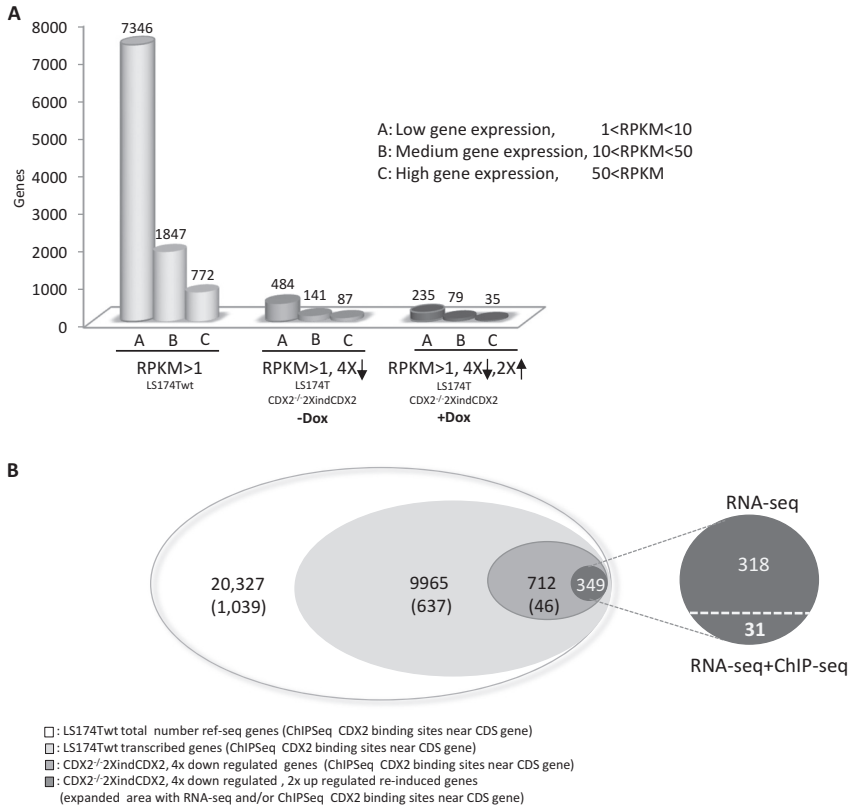


Figure 4. Deep RNA-seq and genome wide ChIP-seq analysis of LS174Twt and induced or uninduced CDX2$^{-/-}$2XindCDX2 PrIITE cells. (**A**) Of the 20 327 human genes included in the analysis around 9965 were found to be expressed with RPKM >1, the majority hereof (7346) were low expressed genes with 1 < RPKM < 10 (A), followed by medium expressed genes (1847) with 10 < RPKM < 50 (**B**) and highly expressed genes (772) with 50 < RPKM (**C**). In uninduced CDX2$^{-/-}$2XindCDX2 (−Dox) PrIITE cells (lacking CDX2), 712 genes expressed in LS174Twt were 4-fold down regulated and after indCDX2 re-induction (+Dox) 349 of these genes were 2-fold upregulated and thus suggested to be CDX2 downstream target genes. (**B**) In order to determine to what degree the genes identified represent direct CDX2 target genes, we included a LS174T ChIP-seq data set and seeked for the presence of potential CDX2 regulatory elements in the vicinity of the RNA-seq suggested CDX2 target genes. Inclusion of this criteria (numbers shown in parenthesis) in addition to the RNA-seq criterion defined above resulted in a dramatic decrease in potential direct CDX2 targets going from 349 to 31 genes. The expanded view at right depicts the number of targets identified using the RNA-seq or RNA-seq+ChIP-seq criteria. In the latter case the targets indicated below the stippled line are listed in Supplementary Table SI.

regulated mucin genes *MUC1* and *MUC7* did not possess any CDX2 binding elements and did not respond to CDX2 changes (Figure 5B and Supplementary Figure S6B). Previously, we have identified CDX2 as a regulator of glycosyltransferase encoded *ST6GALNAC1* expression (45). However, in this study, no CDX2 binding elements were identified in the *ST6GALNAC1* locus (Supplementary Figure S6B), which suggests that CDX2 regulates the expression of this gene by an indirect effect.

The results also show that several known CDX2 targets such as *FUT2* (46), *B3GALT5* (47), *ALP1* (48) and *SI* (28) in LSCDX2 PrIITE cells were unaffected by CDX2 changes (Figure 5B and Supplementary Figure S7) and notably, in the latter three cases these genes were shown not to be expressed in LS174Twt cells. Of notice, the only glycogenes (49) otherwise affected by CDX2 were *FUT3* and *GALNT1* which in the latter case responded inversely to CDX2 changes (Figure 5B).

In order to assess to what degree the global reversibility in gene expression for LSCDX2 PrIITE cells was retained, RPKM values from LS174Twt and LSCDX2 PrIITE cells exposed to different Dox concentrations ranging from 0 to 40 ng/ml were plotted and the regression line for each data calculated (Figure 7 and Supplementary Figure S8). The results show that global reversibility of CDX2 controlled gene expression was profound, and in many cases full reversibility was observed when the lowest 4 ng/ml Dox concentration was used with CDX2$^{-/-}$2XindCDX2 cells. Clearly, the combined RNA-seq and ChIPseq filtering strategy employed improved the specificity in defining true CDX2 regulated genes in our data set (Figure 7C). We speculate, that the observed highest normalization levels of

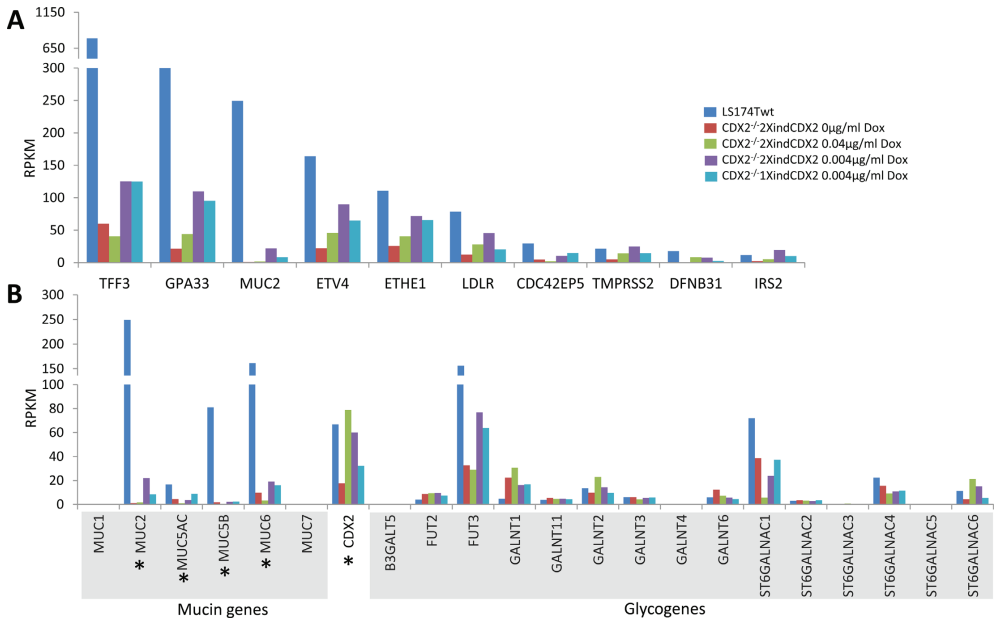


Figure 5. CDX2 regulated target genes. (**A**) Based on the included RNA-seq and ChIPseq criteria (see Figure 4 and Supplementary Table SI) top 10 primary targets downstream of CDX2 with most abundant transcripts (highest RPKM values) were ranked. Well known CDX2 target genes were identified such as *TFF3* and *MUC2*, but novel targets were also identified including *GPA33* and *LDLR*. (**B**) Careful analysis of 208 glycogenes (49) controlling the cellular glycophenotype and 23 mucin genes revealed that *MUC5B*, *MUC6* and *MUC5AC* were controlled by CDX2, whereas *MUC1* and *MUC7* did show no correlation with CDX2 expression. No CDX2 correlation with *FUT2* and *B3GALT5* could be demonstrated, which is in contrast with previous reports. On the other hand, *ST6GALNACI* and -*VI* were shown to correlate with CDX2 expression although no CDX2 binding element could be identified within the vicinity of the CDS of these genes. Interestingly, an inverse correlation was observed for *GALNT1* regulation similar to *SOX2* (see Supplementary Figure S7), with upregulation in the absence of CDX2 expression. Interestingly, the novel CDX2 target *GPA33* gene product has been shown to be post translationally modified by the *GALNT1* encoded enzyme. Genes with ChIP-seq identified CDX2 binding elements in the gene locus vicinity are indicated by an *.

re-induced CDX2 for 4 ng/ml Dox may be due to the role of CDX2 as a molecular rheostat potentially controlling: (i) self-renewal as indicated in Figure 5 and/or (ii) regulated expression of downstream target genes within discrete and narrow CDX2 expression limits (50). Lastly, in order to determine if the genetic manipulations had changed the differentiation state and stemness of CDX2 PrIITE cells generated, we determined the expression profile of 24 marker genes before and after genetic manipulation and/or CDX2 re-induction (Supplementary Figure S7). In all cases (except for *MUC6*), the gene expression profiles of genetically manipulated cells relative to LS174Twt was retained and displayed a distinct intestinal expression pattern. Interestingly, the stemness related gene, *SOX2*, is not expressed in the wt cells, but strongly expressed upon CDX2 KO and clearly decreased upon reinduction. This may suggest that also stemness regulation is CDX2 dependent. The CDX2 targets identified in our data-set is to some extent in agreement with previous findings in a related cellular background (38) (Supplementary Table SI). Of note, the karyotype of the LS174T cells used in our study was determined and found to be in complete agreement with the originally near normal published karyotype (41) (Supplementary Figure S9).

DISCUSSION

Gene regulatory mechanisms are complex and profit from the development of systems in which gene and protein expression is tightly controlled. Tet-regulated expression systems have been widely used for inducible protein expression in mammalian cells. However, current optimized Tet-on systems are still hampered by residual levels of gene expression in the uninduced state in both transiently and stably transfected cells (18,21,22). We hereby demonstrate that the PrIITE system resolves all issues related to leakiness and importantly requires use of tetracycline dosages below cell stress inducing concentrations (24). We first prove the PrIITE concept using a Golgi targeted EYFP model reporter (31). With this tightly controlled inducible reporter system at hand, we were able to demonstrate, to the best of our knowledge for the first time, that leakiness relates to increased expression of the TET3G transactivator element and not increased presence of the inducible GOI, in this case the T2EYFP reporter, Supplementary Figure S10. We furthermore show that stable random integrated pCMV-TET3G transactivator elements causes leaky expression in a substantial sub-population of the cells and we speculate that this sub-population of cells enable uncontrolled Dox induction levels beyond what is achievable in the tightly

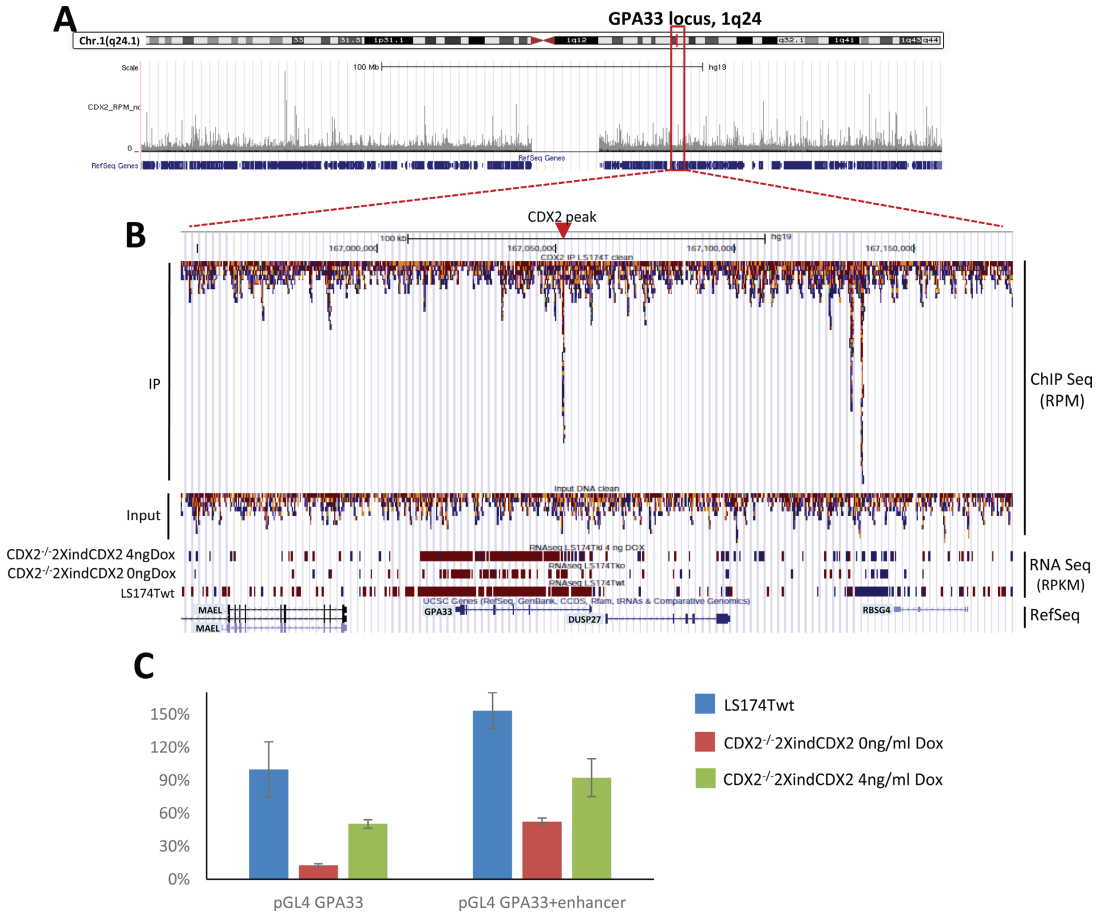


Figure 6. Identification and validation of the novel CDX2 target gene, *GPA33*. (**A**) Panel shows chromosome 1 coverage of CDX2 ChIPseq reads (CDX2IP) and control (input), calculated in 300 bp sliding windows as RPM (reads per million). For each window, input RPM was subtracted from CDX2IP RPM and normalized values were saved as a bedgraph file for visualization in the UCSC Genome Browser. Major CDX2 binding sites (peaks) are shown with indication of the *GPA33* locus framed in red box. (**B**) Expanded display of the GPA33 locus, showing detailed identification of CDX2 ChIPSeq peaks (ChIPseq/RPM) in both LS174T CDX2IP and control input samples. Major *GPA33* CDX2 binding element is indicated by red triangle (CDX2 peak). RNA-seq panel displays the density of individual reads in CDX2$^{-/-}$2XindCDX2 with or without Dox induction (4 or 0 ng respectively) and LS174Twt cells. RefSeq gene panel displays the genomic organization of neighboring genes in the *GPA33* locus. (**C**) *In vitro* promoter analysis of the 0.6 kb CDX2 enhancer element identified in the *GPA33* intron. pGL4-GPA33 and pGL4-*GPA33*enhancer represent the reporter plasmid without or with the enhancer element included. Reporter expression is clearly dependent on both GPA33 CDX2 enhancer and CDX2 presence.

controlled PrIITE cells generated, Supplementary Figure S10B. We next exemplify the potential of the system by clarifying the role of CDX2 in intestinal cells as a consequence of controlled *CDX2* expression. This model system shows that CDX2 predominantly acts as a transcriptional activator and only a limited number of genes seemed to be repressed by CDX2 presence, SOX2 being an example hereof (Supplementary Table SII). We had anticipated that CDX2 targeting in LS174T cells could become problematic due to its proposed essential cellular function and pre-designed our targeting strategy for dealing with essential genes. This was accomplished by incorporating a 'landing pad' in the EPB71 AAVS1 donor integration vector, just downstream of the inducible gene of interest (Supplementary Figure S11). The 'landing pad' encodes the Safe Harbor #1 (SH1) sequence derived from the CHO genome, that has successfully been utilized by us (51) and others (52) for ZFN mediated target integration in CHO cells and thus represents a unique site when integrated in human cells devoid of this sequence. This allows for subsequent donor target integration into the SH1 site. Establishment of a PrIITE cell model for an essential gene can thus be accomplished, by targeting of both the TET3G transactivator and inducible pTRE3G codon optimized GOI transcriptional elements to

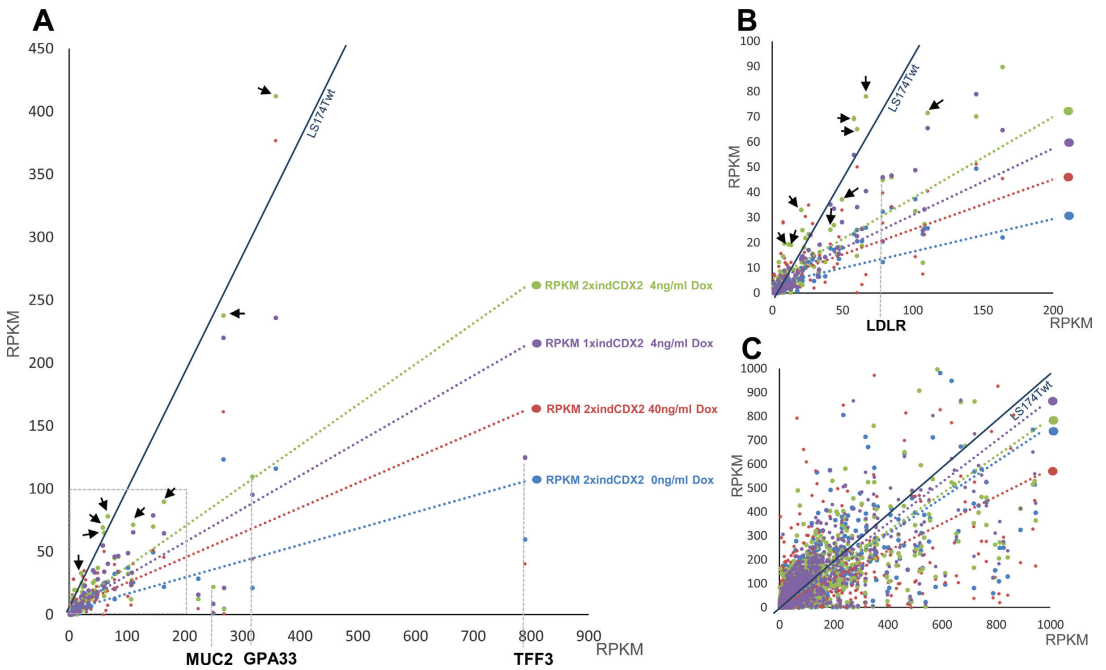


Figure 7. Global Dox induction reversibility of the PrIITE cells relative to LS174Twt cells. In order to include enough data points in the analysis, filtering criteria included all genes with RPKM >1, that were 2× reduced in CDX2$^{-/-}$ KO cells and had a CDX2 binding element/peak in the gene locus. RNA-seq data (RPKM) from wildtype LS174T was plotted against RPKM values from uninduced CDX2$^{-/-}$2XindCDX2 cells without CDX2 expression (blue dotted line, 2XindCDX 0 ng DOX) and under different conditions of CDX2/Dox induction (ng/ml) (green dotted line, 2XindCDX2 4 ng DOX; purple dotted line, 1XindCDX2 4 ng DOX, red dotted line, 2XindCDX2 40 ng DOX). The regression line for each data set is plotted. The dark blue solid line represents the expression levels in LS174Twt cells. (**A**) Global display of RNA-seq RPKM values for the samples analysed. RPKMs for the known CDX2 target genes *MUC2 and TFF3* and the novel target gene, *GPA33*, are indicated by grey dotted vertical lines. Box with gray dotted line is area shown in panel B. Representative examples of genes displaying full reversibility are indicated with arrow heads. (**B**) Zoom from dotted box from panel A with indication of RPKM values for the novel CDX2 target gene, *LDLR*. Representative genes for which complete reversibility was observed are indicated by arrow heads. (**C**) Global display of RNA-seq RPKMs for genes fulfilling the RPKM >1 and 2× reduction filtering criteria but without fulfilling the criteria of a CDX2 binding element/peak in the gene locus. Note the unordered dispersed display of RPKMs relative to the LS174Twt regression line, suggesting that a large proportion of the CDX2 target genes only based on RNA-seq data are not direct targets of CDX2.

the AAVS1 locus, ex. by AAVS1 targeting of the former gene followed by SH1 targeting of the later gene. Hereafter, precise target inactivation of the endogenous gene of interest in the presence of Dox would ensure continued expression of the inducible codon optimized version of the essential GOI. In this project, we did not take advantage of this option, since in the LS174T cellular model CDX2 has not proven to act as a lineage-survival oncogene in contrast to what is observed in other cell line models (40).

Our results also show that LS174T cells have a poor capacity to integrate homology arm flanked donor templates, which is likely attributed to the lacking HR capacity of these and other cell lines used (53–55). In general, precise donor integration remains a challenge in the field and in spite of the fact that CRISPR efficiently induces double stranded breaks, CRISPR mediated donor integration importantly comes at the expense of: (i) 3–20-fold higher indel formation rates at the non-targeted allele (56,57) and (ii) significant off-target/random integration frequencies (58). These concerns were evident in a recent *in vivo* study based on CRISPR target integration of inducible TRE3G promoter elements, where 'leaky' expression in 33% of clones analyzed was identified (59). In light of these CRISPR related issues, the ObLiGaRe targeting strategy by Marcello Maresca *et al.* remains to be an efficient and reliable target integration strategy for most cells (60–63), including HR compromised cells such as CHO cells (64,65) and colorectal cancer lines as described in this study and by Marcello Maresca et al.

A whole set of novel genes, including *GPA33* (44), *LDLR* (66), *MUC6*, *MUC5B* and *MUC5AC* were identified as CDX2 targets. To our knowledge, this has not previously been demonstrated and suggests *CDX2* as a locus control gene of the 11p15 mucin gene cluster (67–69) that includes *MUC2*, *MUC6*, *MUC5B* and *MUC5AC*. In this respect, a recent study of 295 gastric adenocarcinomas has shown that *GPA33*, *MUC6* and *MUC2* were among the 10 most abundant differentially regulated genes in 295 analyzed microsatellite instable (MSI) colorectal adenocarcinomas (70). Although *CDX2* profiling was not evident in this study it is stimulating to assume that MSI has impacted on *CDX2* expression in the tumors studied. Furthermore, *GPA33* is

expressed in cancer of the gastrointestinal tract, namely in over 95% of human colon cancers and has thus been suggested as an attractive novel colorectal carcinoma therapeutic target (71,72). Of particular notice, *GPA33* gene product has been shown to be post translationally modified by the *GALNT1* encoded GalNAc-transferase 1 enzyme (73,74), that we here show is inversely influenced by CDX2 (Figure 5B). Our finding that *LDLR* is a major direct CDX2 target gene is surprising since *LDLR* has been primarily found to be regulated by sterol regulated element-binding proteins (SREBPs) (75,76). We speculate that CDX2 plays novel roles in the co-regulated appearance of these novel gene targets.

In summary, we hereby describe a novel approach for generating an inducible 'non leaky' isogenic knockout-rescue system. The principle is based on precise genome integration of defined copies of third generation Tet-On elements, which enables for an 'all or none' context for studying the mechanism by which a given gene works in a cell. We exemplify the utility of the system by establishing a colonic CDX2 PrIITE cell system that uncovered novel molecular interactions governed by CDX2.

MATERIALS & CORRESPONDENCE

All reagents and cell lines used in the study are available upon request for research purposes under a material transfer agreement, except for TRE3G and TET3G containing plasmids, due to restrictions over distribution of plasmids containing the TRE3G promoter or TET3G ORF. The RNA-seq and ChIP-seq data shown has been uploaded at the NIH GEO server: http://www.ncbi.nlm.nih.gov/geo/ and is accessible using the following accession number GSE97273. Plasmids described are available from Addgene (https://www.addgene.org/). All correspondence should be addressed to Eric P. Bennett.

NOTE ADDED IN PROOF

Extended experiments conducted during proofs of this manuscript revealed that the GPA33 promoter described in Figure 6 possess bi-directional transcriptional activity, likely mediated through presence of multiple LINEs contained within the 1KBp GPA33 promoter analyzed. LINEs have been shown to mediate bi-directional expression (Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. and Myers, R.M. (2004) An abundance of bidirectional promoters in the human genome. Genome Res.14:62–66; Core, L.J., Waterfall, J.J. and Core, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 322:1845–1848) potentially affecting 4% of all human genes (Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol. Cell. Biol. 2001, 1973–1985; Criscione, S.W., Theodosakis, N., Micevic, G., Cornish, T.C., Burns, K.H., Neretti, N., and Rodić, N. (2016) Genome-wide characterization of humanL1 antisense promoter-driven transcripts. BMC Genomics, 17:463).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Camilla Andersen, Claire Katrine Beier Holden, Asli Silahtaroglu, Elisabeth Larsen, Mads Bak (Departments of Odontology and Cellular and Molecular Medicine, Copenhagen Center for Glycomics, University of Copenhagen) and Katja Dahlgaard, (Department of Science and Environment, Roskilde University, Denmark) for excellent technical assistance. We thank Kim Madsen at CLC Bioscience for guidance and access to CLC Workbench.

Author contributions: R.P. planned and performed all ZFN targeting, IHC, FACS and live imaging experiments. L.H., M.C and C.M. planned and performed RNA-seq and ChIPSeq experiments. S.L. and J.D. performed all *in vitro* promoter analysis. J.B.H. performed Dox induction and FACS experiments. J.T.T., R.A. and L.D. contributed with writing parts of the manuscript. E.P.B. designed and planned all experiments and wrote the manuscript.

FUNDING

University of Copenhagen Excellence Programme for Interdisciplinary Research [CDO2016]; The Danish Council for Independent Research [4004-00140B], Novo Nordisk Foundation, Danish National Research Foundation [DNRF107], FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação [POCI-01-0145-FEDER-007274, POCI-01-0145-FEDER-016390]. Funding for open access charge: Danish National Research Foundation [DNRF107].

Conflict of interest statement. None declared.

REFERENCES

1. McClintock,B. (1950) the origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.*, **36**, 344–355.
2. Goth,R. and Rajewsky,M.F. (1972) Ethylation of nucleic acids by ethylnitrosourea-l-14C fetal and adult Rat1. *Cancer Res.*, **32**, 1501–1505.
3. Hinton,C. and Whittinghill,M. (1950) The distribution of X-ray induced crossovers from curly inversion heterozygotes of Drosophila melanogaster females. *Proc. Natl. Acad. Sci. U.S.A.*, **36**, 552–558.
4. Capecchi,M.R. (1980) High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell*, **22**, 479–488.
5. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
6. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.a. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
7. Geurts,A.M., Cost,G.J., Freyvert,Y., Zeitler,B., Jeffrey,C., Choi,V.M., Jenkins,S.S., Wood,A., Cui,X., Meng,X. *et al.* (2010) Knockout rats produced using designed zinc finger nucleases. *Science*, **325**, 2009–2011.
8. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by

double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806–811.
9. Liu,J., Deutsch,U., Fung,I. and Lobe,C.G. (2014) Conditional and inducible transgene expression in endothelial and hematopoietic cells using Cre/loxP and tetracycline-off systems. *Exp. Ther. Med.*, **8**, 1351–1356.
10. Kuhn,R., Schwenk,F., Aguet,M. and Rajewsky,K. (1995) Inducible gene targeting in mice. *Science*, **269**, 1427–1428.
11. Braselmann,S., Graninger,P. and Busslinger,M. (1993) A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 1657–1661.
12. Kennedy,M.J., Hughes,R.M., Peteya,L.A., Schwartz,J.W., Ehlers,M.D. and Tucker,C.L. (2010) Rapid blue-light – mediated induction of protein interactions in living cells. *Nat. Methods*, **7**, 12–16.
13. No,D., Yao,T.P. and Evans,R.M. (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 3346–3351.
14. Unger,B. and Hillen,W. (1984) Nucleotide sequence of the repressor gene of the RAI tetracycline resistance determinant: structural and functional comparison with three related Tet repressor genes Bernhard. *Nucleic Acids Res.*, **12**, 7693–7703.
15. Postle,K., Nguyen,T.T. and Bertrand,K.P. (1984) Nucleic acids research. *Nucleic Acids Res.*, **12**, 4849–4863.
16. Gossen,M. and Bujard,H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 5547–5551.
17. Gossen,M., Freundlieb,S., Bender,G., Müller,G., Hillen,W. and Bujard,H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science*, **268**, 1766–1769.
18. Das,A.T., Zhou,X., Metz,S.W., Vink,M.A. and Berkhout,B. (2016) Selecting the optimal Tet-On system for doxycycline-inducible gene expression in transiently transfected and stably transduced mammalian cells. *Biotechnol. J.*, **11**, 71–79.
19. Urlinger,S., Baron,U., Thellmann,M., Hasan,M.T., Bujard,H. and Hillen,W. (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7963–7968.
20. Freundlieb,S., Schirra-Müller,C. and Bujard,H. (1999) A tetracycline controlled activation/repression system with increased potential for gene transfer into mammalian cells. *J. Gene Med.*, **1**, 4–12.
21. Loew,R., Heinz,N., Hampf,M., Bujard,H. and Gossen,M. (2010) Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.*, **10**, 81.
22. Heinz,N., Schambach,A., Galla,M., Maetzig,T., Baum,C., Loew,R. and Schiedlmeier,B. (2011) Retroviral and transposon-based tet-regulated all-in-one vectors with reduced background expression and improved dynamic range. *Hum. Gene Ther.*, **22**, 166–176.
23. Das,A.T., Liliane,T. and Berkhout,B. (2016) Tet-On systems for doxycycline-inducible gene expression. *Curr. Gene Ther.*, **16**, 156–167.
24. Moullan,N., Mouchiroud,L., Houtkooper,R.H., Auwerx,J., Moullan,N., Mouchiroud,L., Wang,X., Ryu,D., Williams,E.G. and Mottis,A. (2015) Tetracyclines disturb mitochondrial function across eukaryotic models: a call for caution in biomedical article tetracyclines disturb mitochondrial function across eukaryotic models: a call for caution in biomedical research. *Cell Rep.*, **10**, 1681–1691.
25. Bensinger,S.J., Graeber,T.G. and Christofk,H.R. (2013) Doxycycline alters metabolism and proliferation of human cell lines. *PLoS One*, **8**, 8–14.
26. Brüning,A., Brem,G.J., Vogel,M. and Mylonas,I. (2013) Tetracyclines cause cell stress-dependent ATF4 activation and mTOR inhibition. *Exp. Cell Res.*, **320**, 281–289.
27. Zhou,X., Vink,M., Klaver,B., Berkhout,B. and Das,A.T. (2006) Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther.*, **13**, 1382–1390.
28. Suh,E., Chen,L., Taylor,J. and Traber,P.G. (1994) A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol. Cell. Biol.*, **14**, 7340–7351.
29. Duda,K., Lonowski,L.a., Kofoed-Nielsen,M., Ibarra,A., Delay,C.M., Kang,Q., Yang,Z., Pruett-Miller,S.M., Bennett,E.P., Wandall,H.H. *et al.* (2014) High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res.*, doi:10.1093/nar/gku251.
30. Maresca,M., Lin,V.G., Guo,N. and Yang,Y. (2013) Obligate Ligation-Gated Recombination (ObLiGaRe): custom-designed nucleases mediated targeted integration through nonhomologous end joining. *Genome Res.*, doi:10.1101/gr.145441.112.
31. Wandall,H.H., Rumjantseva,V., Sørensen,A.L.T., Patel-Hett,S., Josefsson,E.C., Bennett,E.P., Italiono,J.E., Henrik,C., H,H.J. and Hoffmeister,K.M. (2012) The origin and function of platelet glycosyltransferases Glycosyltransferase assays Immunoblot analysis Flow cytometric analysis of platelet glycosyltransferases. *Blood*, **120**, 626–635.
32. Yang,Z., Steentoft,C., Hauge,C., Hansen,L., Thomsen,a. L., Niola,F., Vester-Christensen,M.B., Frodin,M., Clausen,H., Wandall,H.H. *et al.* (2015) Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.*, **43**, e59.
33. Lonowski,L.A., Narimatsu,Y., Riaz,A., Delay,C.E., Yang,Z., Niola,F., Duda,K., Ober,E.A., Clausen,H., Wandall,H.H. *et al.* (2017) Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat. Protoc.*, **12**, 581–603.
34. Coskun,M., Troelsen,J.T. and Nielsen,O.H. (2011) The role of CDX2 in intestinal homeostasis and inflammation. *Biochim. Biophys. Acta*, **1812**, 283–289.
35. Ji,H., Jiang,H., Ma,W., Johnson,D.S., Myers,R.M. and Wong,W.H. (2008) An integrated system CisGenome for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, **26**, 1293–1300.
36. Gao,N., White,P. and Kaestner,K.H. (2010) Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev. Cell*, **16**, 588–599.
37. Mesquita,P., Jonckheere,N., Almeida,R., Ducourouble,M.P., Serpa,J., Silva,E., Pigny,P., Santos Silva,F., Reis,C., Silberg,D. *et al.* (2003) Human MUC2 mucin gene is transcriptionally regulated by Cdx homeodomain proteins in gastrointestinal carcinoma cell lines. *J. Biol. Chem.*, **278**, 51549–51556.
38. Boyd,M., Hansen,M., Jensen,T.G.K., Perearnau,A., Olsen,A.K., Bram,L.L., Bak,M., Tommerup,N., Olsen,J. and Troelsen,J.T. (2010) Genome-wide analysis of CDX2 binding in intestinal epithelial cells (Caco-2). *J. Biol. Chem.*, **285**, 25115–25125.
39. Bennett,E.P., Mandel,U., Clausen,H., Gerken,T.a, Fritz,T.a and Tabak,L.a (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, **22**, 736–756.
40. Salari,K., Spulak,M.E., Cuff,J., Forster,A.D., Giacomini,C.P. and Huang,S. (2012) CDX2 is an amplified lineage-survival oncogene in colorectal cancer Keyan. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 3196–3205.
41. Rutzky,L.P., Kaye,C.I., Siciliano,M.J., Chao,M. and Kahan,B.D. (1980) Longitudinal karyotype and genetic signature analysis of cultured human colon adenocarcinoma cell lines LS180 and LS174T. *Cancer Res.*, **40**, 1443–1448.
42. Göttke,M.S., Keller,K., Belley,A., Garcia,R.-M., Hollingsworth,M.A., Mack,D.R. and Chadee,K. (1998) Functional heterogeneity of colonic adenocarcinoma mucins for inhibition of Entamoeba histolytica adherence to target cells. *J. Eukaryot. Microbiol.*, **45**, 17S–23S.
43. Bu,X.D., Li,N., Tian,X.Q. and Huang,P.L. (2011) Caco-2 and LS174T cell lines provide different models for studying mucin expression in colon cancer. *Tissue Cell*, **43**, 201–206.
44. Heath,J.K., White,S.J., Johnstone,C.N., Catimel,B., Simpson,R.J., Moritz,R.L., Tu,G.F., Ji,H., Whitehead,R.H., Groenen,L.C. *et al.* (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 469–474.
45. Pinto,R., Barros,R., Pereira-Castro,I., Mesquita,P., da Costa,L.T., Bennett,E.P., Almeida,R. and David,L. (2015) CDX2 homeoprotein is involved in the regulation of ST6GalNAc-I gene in intestinal metaplasia. *Lab. Invest.*, **95**, 1–10.
46. Sakuma,K., Aoki,M. and Kannagi,R. (2012) Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF / bFGF-induced epithelial—mesenchymal transition. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 7776–7781.

47. Isshiki,S., Kudo,T., Nishihara,S., Ikehara,Y., Togayachi,A., Furuya,A., Shitara,K., Kubota,T., Watanabe,M., Kitajima,M. *et al.* (2003) Lewis type 1 antigen synthase (β3Gal-T5) is transcriptionally regulated by homeoproteins. *J. Biol. Chem.*, **278**, 36611–36620.
48. Mitchelmore,C., Troelsen,J.T., Spodsberg,N., Sjöström,H. and Norén,O. (2000) Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. *Biochem. J.*, **346**, 529–535.
49. Hansen,L., Lind-thomsen,A., Joshi,H.J., Pedersen,N.B., Have,C.T., Kong,Y., Wang,S., Sparso,T., Grarup,N., Vester-christensen,M.B. *et al.* (2015) A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology*, **25**, 211–224.
50. Rizzino,A. (2009) Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **1**, 228–236.
51. Yang,Z., Wang,S., Halim,A., Schulz,M.A., Frodin,M., Rahman,S.H., Vester-Christensen,M.B., Behrens,C., Kristensen,C., Vakhrushev,S.Y. *et al.* (2015) Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.*, doi:10.1038/nbt.3280.
52. Bahr,S., Cortner,L., Ladley,S. and Borgschulte,T. (2013) Evaluating the effect of chromosomal context on zinc finger nuclease efficiency. *BMC Proc.*, **7**, P3.
53. Mansour,W.Y., Schumacher,S., Rosskopf,R., Rhein,T., Schmidt-Petersen,F., Gatzemeier,F., Haag,F., Borgmann,K., Willers,H. and Dahm-Daphi,J. (2008) Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. *Nucleic Acids Res.*, **36**, 4088–4098.
54. Cristea,S., Freyvert,Y., Santiago,Y., Holmes,M.C., Urnov,F.D., Gregory,P.D. and Cost,G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.*, **110**, 871–880.
55. Yang,Z., Wang,S., Halim,A., Schulz,M.A., Frodin,M., Rahman,S.H., Vester-christensen,M.B., Behrens,C., Kristensen,C., Vakhrushev,S.Y. *et al.* (2015) Engineered CHO cells for production of diverse , homogeneous glycoproteins. *Nat. Biotechnol.*, **33**, 842–844.
56. Inui,M., Miyado,M., Igarashi,M., Tamano,M., Kubo,A., Yamashita,S., Asahara,H., Fukami,M. and Takada,S. (2014) Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci. Rep.*, **4**, 5396.
57. Lee,J.S., Grav,L.M., Pedersen,L.E., Lee,G.M. and Kildegaard,H.F. (2016) Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. Bioeng.*, **9999**, 1–17.
58. He,X., Tan,C., Wang,F., Wang,Y., Zhou,R., Cui,D., You,W., Zhao,H., Ren,J. and Feng,B. (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.*, **44**, e85.
59. Dow,L.E., Fisher,J., O'Rourke,K.P., Muley,A., Kastenhuber,E.R., Livshits,G., Tschaharganeh,D.F., Socci,N.D. and Lowe,S.W. (2015) Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.*, **33**, 390–394.
60. Wang,J., Exline,C.M., DeClercq,J.J., Llewellyn,G.N., Hayward,S.B., Li,P.W.-L., Shivak,D.A., Surosky,R.T., Gregory,P.D., Holmes,M.C. *et al.* (2015) Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat. Biotechnol.*, **33**, 1256–1263.
61. Cai,Y., Laustsen,A., Zhou,Y., Sun,C., Anderson,M.V., Li,S., Uldbjerg,N., Luo,Y., Jakobsen,M.R. and Mikkelsen,J.G. (2016) Targeted, homology-driven gene insertion in stem cells by ZFN-loaded ' all-in-one ' lentiviral vectors. *Elife*, e12213.
62. Coluccio,A., Miselli,F., Lombardo,A., Marconi,A., Malagoli Tagliazucchi,G., Gonçalves,M.a, Pincelli,C., Maruggi,G., Del Rio,M., Naldini,L. *et al.* (2013) Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther.*, **21**, 1695–1704.
63. Moehle,E.a, Rock,J.M., Lee,Y.-L., Jouvenot,Y., DeKelver,R.C., Gregory,P.D., Urnov,F.D. and Holmes,M.C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3055–3060.
64. Yamamoto,Y., Bliss,J. and Gerbi,S.A. (2015) Whole organism genome editing: targeted large DNA insertion via ObLiGaRe nonhomologous end-joining in vivo capture. *G3 (Bethesda).*, **5**, 1843–1847.
65. Yang,Z., Wang,S., Halim,A., Schulz,M.A., Frodin,M., Rahman,S.H., Vester-Christensen,M.B., Behrens,C., Kristensen,C., Vakhrushev,S.Y. *et al.* (2015) Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.*, **33**, 2014–2017.
66. Brown,M.S., Goldstein,J.L., Arnold,W. and Perdue,S. (1974) Expression of the familial hypercholesterolemia gene in heterozygotes: mechanism for a dominant disorder in man. *Science*, **185**, 61–63.
67. Van Cong,N., Aubert,J.P., Gross,M.S., Porchet,N., Degand,P. and Frézal,J. (1990) Assignment of human tracheobronchial mucin gene(s) to 11p15 and a tracheobronchial mucin-related sequence to chromosome 13. *Hum. Genet.*, **86**, 167–172.
68. Gerard,C., Eddy,R.L. and Shows,T.B. (1990) The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat structure: evidence for a common mucin in airway and gastrointestinal tissue. *J. Clin. Invest.*, **86**, 1921–1927.
69. Velcich,A., Palumbo,L., Selleri,L., Evans,G. and Augenlicht,L. (1997) Organization and regulatory aspects of the human intestinal mucin gene (MUC2) locus. *J. Biol. Chem.*, **272**, 7968–7976.
70. Wheeler,D.a., Gibbs,R.a., Muzny,D.M., Bainbridge,M.N., Chang,K., Dinh,H.H., Drummond,J.a., Fowler,G., Kovar,C.L., Lewis,L.R. *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**, 330–337.
71. Sakamoto,J., Kojima,H., Kato,J., Hamashima,H. and Suzuki,H. (2000) Organ-specific expression of the intestinal epithelium-related antigen A33, a cell surface target for antibody-based imaging and treatment in gastrointestinal cancer. *Cancer Chemother. Pharmacol.*, **46**(Suppl), S27–S32.
72. Ciprotti,M., Chong,G., Gan,H.K., Chan,A., Murone,C., MacGregor,D., Lee,F.-T., Johns,T.G., Heath,J.K., Ernst,M. *et al.* (2014) Quantitative intratumoural microdistribution and kinetics of (131)I-huA33 antibody in patients with colorectal carcinoma. *EJNMMI Res.*, **4**, 1–10.
73. Steentoft,C., Vakhrushev,S.Y., Joshi,H.J., Kong,Y., Vester-Christensen,M.B., Schjoldager,K.T.-B.G., Lavrsen,K., Dabelsteen,S., Pedersen,N.B., Marcos-Silva,L. *et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.*, **32**, 1478–1488.
74. Kong,Y., Joshi,H.J., Schjoldager,K.T.-B.G., Madsen,T.D., Gerken,T.a, Vester-Christensen,M.B., Wandall,H.H., Bennett,E.P., Levery,S.B., Vakhrushev,S.Y. *et al.* (2014) Probing polypeptide GalNAc-transferase isoform substrate specificities by in vitro analysis. *Glycobiology*, **0**, 1–11.
75. Horton,J.D., Goldstein,J.L. and Brown,M.S. (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.*, **109**, 1125–1131.
76. Shimano,H. (2001) Sterol regulatory element-binding proteins (SREBPs): transcriptional regulators of lipid synthetic genes. *Prog. Lipid Res.*, **40**, 439–452.

Supplementary material:

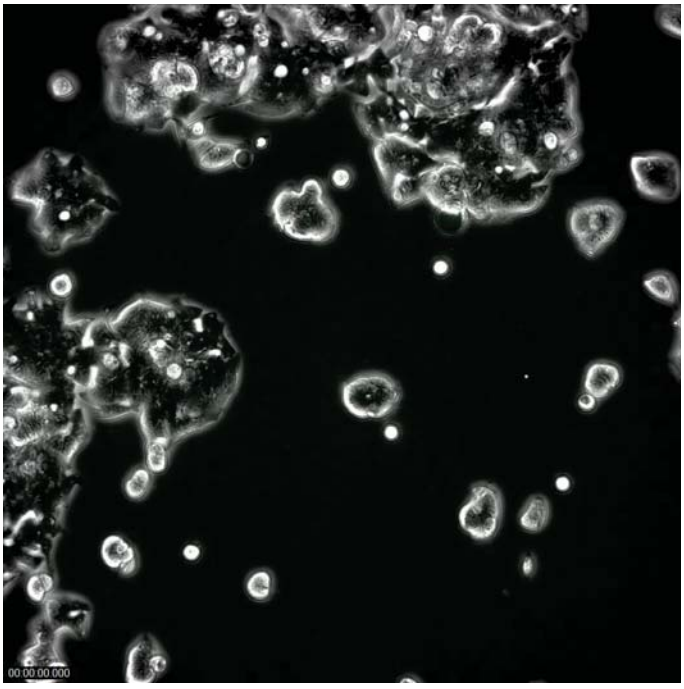
Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression.

Rita Pinto, Lars Hansen, John Hintze, Raquel Almeida, Sylvester Larsen, Mehmet Coskun, Johanne Davidsen, Cathy Mitchelmore, Leonor David, Jesper Thorvald Troelsen and Eric Paul Bennett

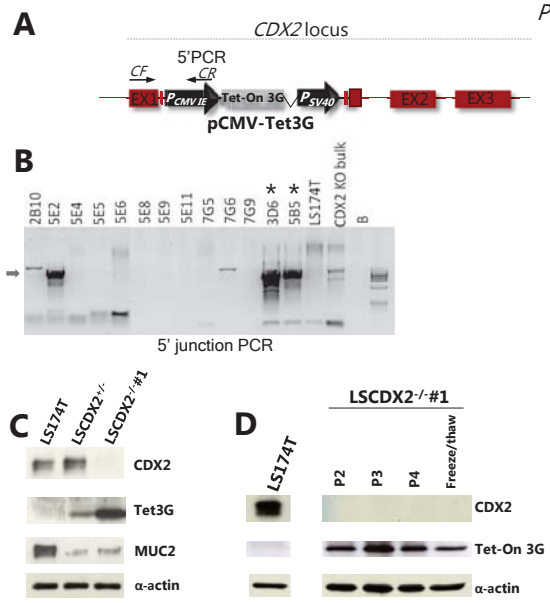
Supplementary material contains:

Supplementary video material
Supplementary Figures 1-11
Supplementary Tables I-IV

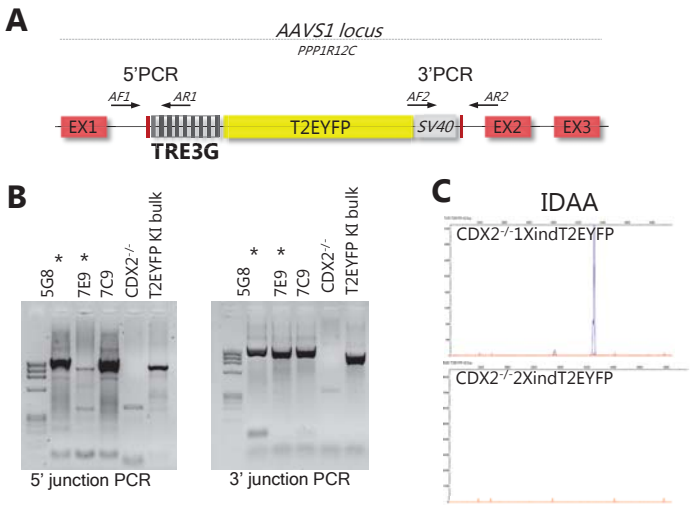
LS174T T2EYFP PrIITE Cells



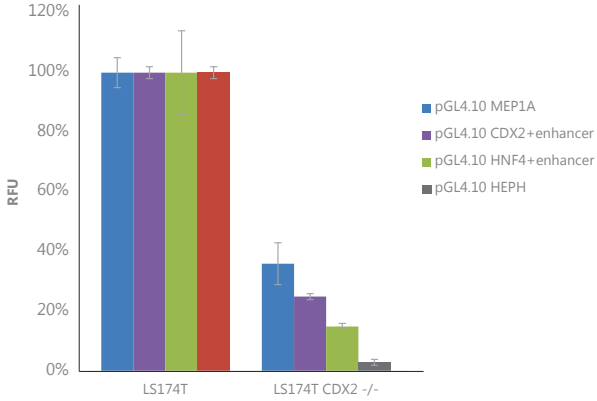
Supplementary video material
24h Time-lapse video showing T2EYFP expression in CDX2$^{-/-}$2XindT2EYFP clone upon induction with 0.5µg/mL of doxycycline.



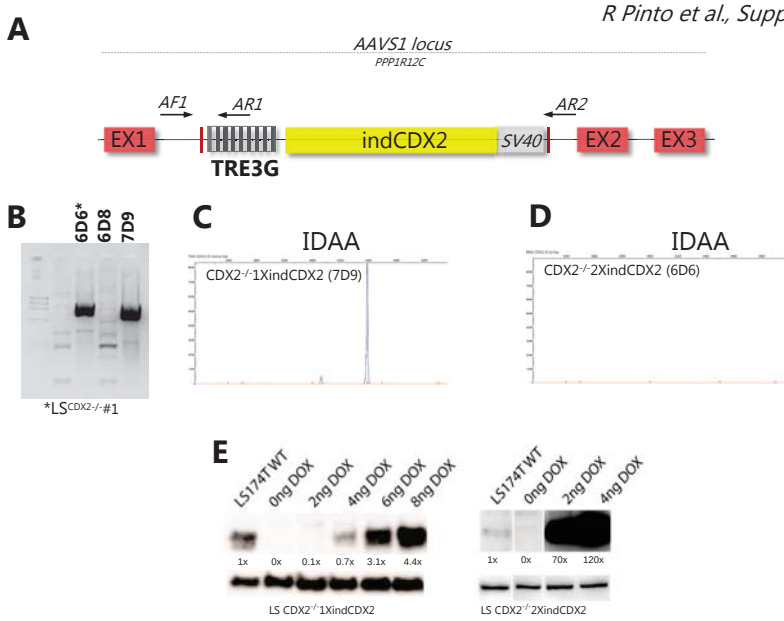
Supplementary Figure 1
Target integration of tetracycline responsive transactivator (TET3G) donor elements into *CDX2* exon 1. **A.** Schematic diagram of the pCMV-Tet3G targeting construct, with the regulatory elements controlling constitutive Tet3G expression, targeted to *CDX2* exon1. Forward and reverse primers used for determining proper target integration at the *CDX2* locus are shown (CF and CR respectively). **B.** Correct integration was detected by successful amplification of a 1.3 kb product (5'-junction PCR). Furthermore, amplification with primers flanking the *CDX2* ZFN target site would detect wild type (wt) allele presence, while lack of successful amplification would indicate correct targeting due to amplification settings chosen, not allowing for amplification across the integrated donor target sequence (data not shown). Asterix indicates clones included in this study. LS174T and CDX2KO bulk represent untransfected and 2 days post electroporation transfected cells respectively. No positive signal could be detected on bulk by target integration using homologous recombination and 800bp homology flanking donor arms on donor construct, data not shown. **C.** Western blot detection of CDX2, Tet3G and MUC2 in LS174Twt, CDX2$^{+/-}$ or CDX2$^{-/-}$ cells. Monoclonals against CDX2, Tet3G, MUC2 and α-actin (control) show, that CDX2 protein is only expressed in wt and CDX2$^{+/-}$ cells and absent in CDX2$^{-/-}$ cells. 104 clones were analyzed in total and copy number dependent Tet3G expression was detected in the respective PrIITE cells and MUC2 protein was clearly decrease in expression in the cells shown. α-actin control is shown below. **D.** Western blot analysis of LS174Twt cells and bi-allelic targeted 3D6 clone (CDX2$^{-/-}$) in the latter case after prolonged passage (P2, P3 and P4) and freeze/thawing cycle. Stable Tet3G expression was observed and absence of *CDX2* maintained.



Supplementary Figure 2
Target integration of T2EYFP "proof of principle" reporter donor elements to *AAVS1* intron 1. **A.** Schematic illustration of *p*AAVS1-TRE3G-T2EYFP targeted to AAVS1 intron 1 locus. Position of the genetic primers used for determining correct targeting are shown with arrows above the Tet3G responsive elements/TRE3G promoter and SV40 3'UTR, AF1/2 and AR1/2. **B.** Junction PCR showing correct integration of donor construct at the *AAVS1* target site. Asterix indicates clones included in this study. CDX2$^{-/-}$ and T2EYFP KI bulk represent untransfected and 2day post transfection controls. **C.** Detection of wild type allele by IDAA assay[50]. Presence of detected amplicon/peak represents untargeted wild type allele while lack of peak represents bi-allelic target integration of *p*AAVS1-TRE3G-T2EYFP donor construct.

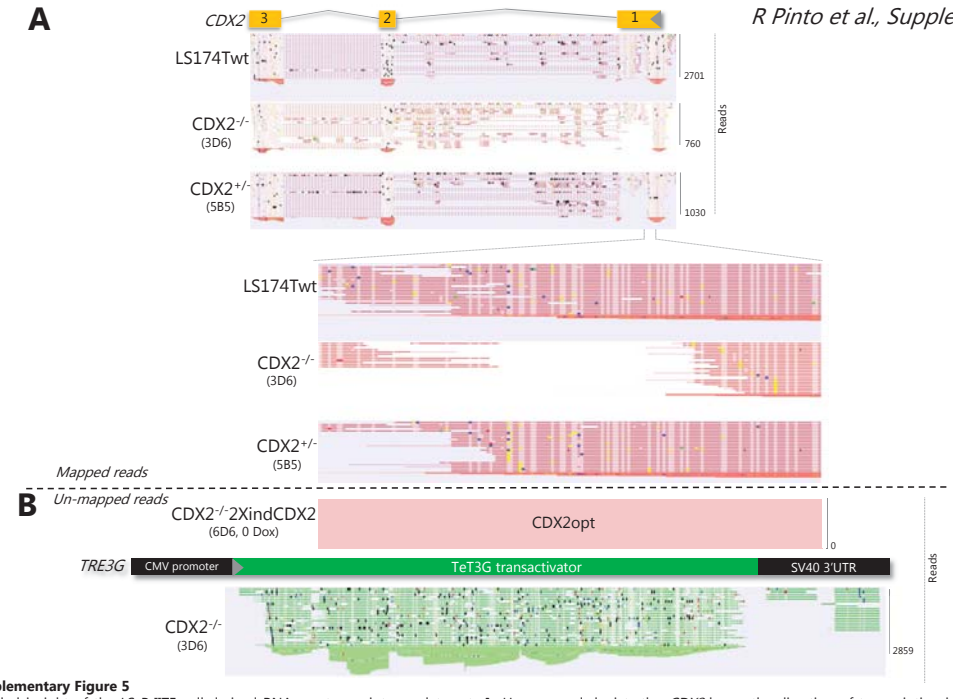


Supplementary Figure 3
Validating the lack of functional CDX2 expression in the CDX2$^{-/-}$ 3D6 clone. Using known CDX2 dependent enhancer elements[30], promoter analysis was conducted in LS174T cells and 3D6 cells and the relative fluorescence units normalized to levels observed in LS174Twt cells. Most prominent >95% repressed expression was observed for the HEPH reporter, clearly demonstrating the lack of functional CDX2 in the 3D6 cells.



Supplementary Figure 4

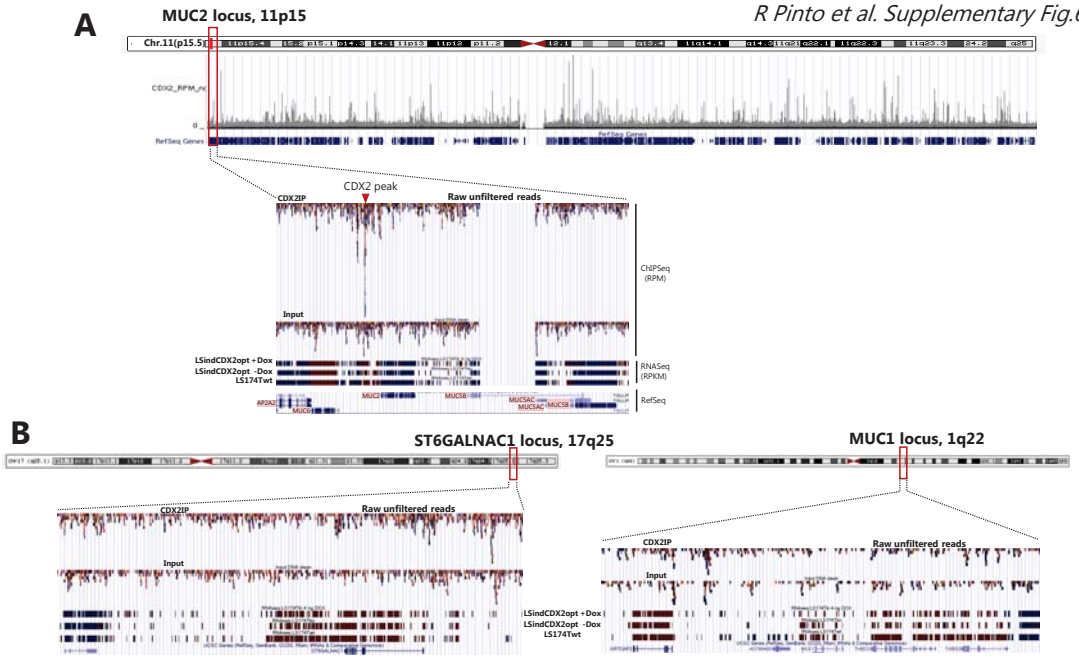
Target integration of Tet3G inducible codon optimized CDX2 donor elements to *AAVS1* intron 1. **A.** Schematic illustration of the *p*AAVS1-TRE3G-CDX2opt donor construct targeted to *AAVS1* intron 1. Position of the genetic primers used for determining correct targeting are shown with arrows above the Tet3G responsive element/TRE3G promoter and flanking the SV40 3'UTR, AF1, AR1/2. **B.** Junction PCR showing correct target integration in CDX2$^{-/-}$ cells (LS$^{CDX2-/-}$#1). Clone 6D6 was used for further analyses in this study, indicated by an asterix. **C.** Detection of wild type *AAVS1* allele by IDAA assay[50]. Presence of detected amplicon/peak represents untargeted wild type allele presence and thus integration of only one copy of *p*AAVS1-TRE3G-CDX2opt donor construct (CDX2$^{-/-}$1XindCDX2 (7D9)). **D.** Lack of peak represents bi-allelic target integration and thus integration of 2 copies of *p*AAVS1-TRE3G-CDX2opt donor construct (CDX2$^{-/-}$2XindCDX2 (6D6)). **E.** Western blot analysis for CDX2 expression in LS174Twt and PrIITE cells (LS CDX2$^{-/-}$1XindCDX2 or LS CDX2$^{-/-2}$XindCDX2) with or without variable Dox concentrations. Equal amounts of total protein (10µg) were loaded on the primary CDX2-88 anti-body probed Westerns shown. Lower panels illustrate loading control probings with anti GAPDH (Fitzgerald Industries International, USA, cat#10R-G109a) or anti Vinculin (Abcam, EPR8185) respectively. For clarity results were quantified by densitometry using ImageJ 1.51j8 software (https://imagej.nih.gov/ij/download.html) and presented as fold difference relative to CDX2 levels detected in LS174Twt cells.



R Pinto et al., Supplementary Fig. 5

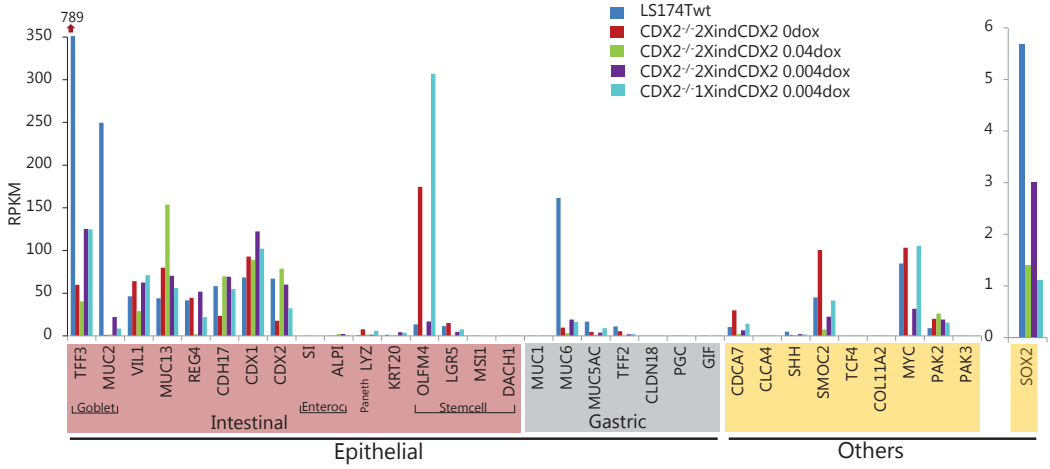
Supplementary Figure 5

Detailed insight of the LS-PrIITE cell derived RNA-seq transcriptome data set. **A.** Upper panel depicts the *CDX2* locus, the direction of transcription is indicated by an arrow. Sequence fragments/read densities for LS174Twt, *CDX2*$^{-/-}$ (3D6) and *CDX2*$^{+/-}$ (5B5) are shown across the locus as red lines and below each RNA-seq panel the exon read densities are indicated with red profiles in the ligth blue shaded panels. The total number of locus reads is indicated to the right of each panel. An expanded view of the pCMV-Tet3G donor integration targeted site within *CDX2* exon 1 is shown below the three upper RNA-seq panels. Note the lack of reads in the CDX2$^{-/-}$ 3D6 clone at the exact sites of biallelic pCMV-Tet3G integration and that in spite of this, *CDX2* locus specific transcripts are detectable but at ≈25% abundance relative to LS174Twt levels. **B.** Detailed insight into un-mapped reads of RNAseq sequences that could not be matched to sequences in the hg19 database. Since codon optimized *CDX2* and *Tet3G* transactivator encoded donor integration elements were used, transcripts from these respective transcription units were not identified in the hg19 mapped reads data set. Upper empty light blue panel depicts the complete absence of inducible codon optimized CDX2 transcripts in uninduced CDX2$^{-/-}$2XindCDX2 6D6 cells. Lower panel depicts the abundant reads identified for the constitutively expressed Tet3G transactivator in CDX2$^{-/-}$ 3D6 cells that found the basis for most of the cells included in this study. Number of reads are indicated to the right of the panels.



Supplementary Figure 6

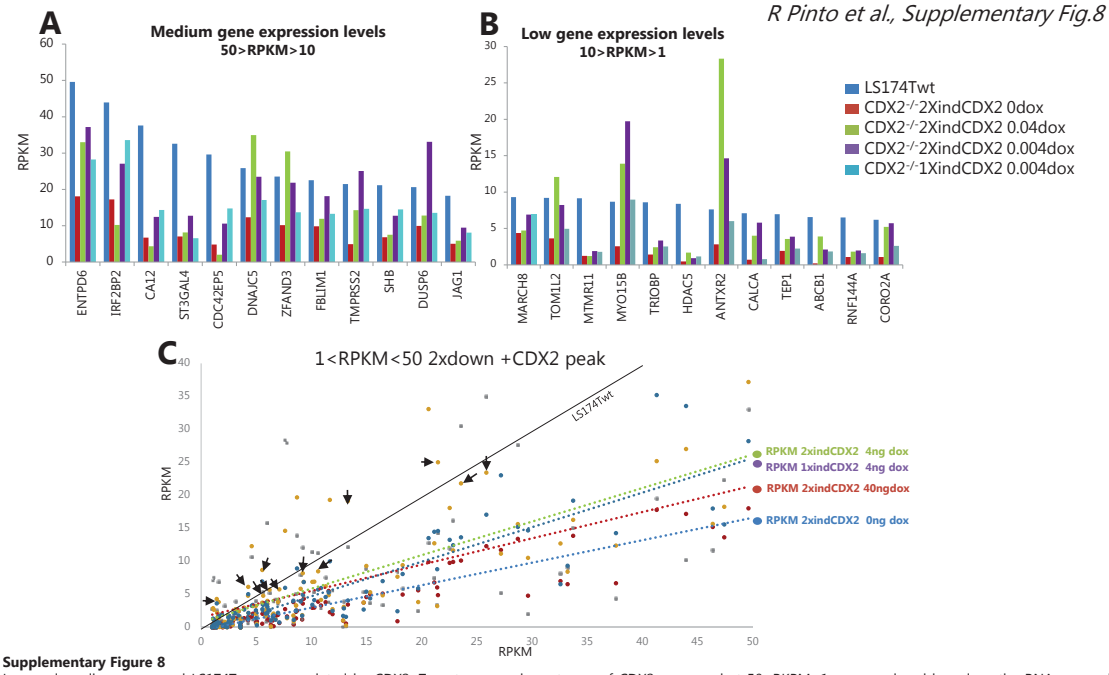
RNA-seq and ChIPSeq identified primary CDX2 target genes. **A.** Panel shows chromosome 11 coverage of CDX2IP (CDX2 ChIPseq) reads. Coverage was determined as described in **Fig.6**. Major CDX2 binding sites (peaks) are shown with indication of the *MUC2* 11p15 mucin gene locus framed in red box. Expanded display of the 11p15 mucin gene locus, showing detailed identification of CDX2 ChIP-Seq (CDX2IP) and control (Input) peaks in both LS174Twt samples. The major *MUC2* CDX2 binding element is indicated by red triangle (CDX2 peak). Lower RNA-seq panels display the density of individual reads in CDX2$^{-/-}$2XindCDX2 cells with or without Dox induction (4ng or 0ng respectively) and LS174Twt cells. RefSeq gene panel displays the genomic organization of genes within the 11p15 locus. Notably, the major peak (CDX2 enhancer element) was positioned in the intergenic *MUC6* and *MUC2* region. The deleted region shown in panel is due to an unannotated region in the hg19 genome reference data base used for the analysis. **B.** Chromosome 17 and 1 diplays with zoom in panels shown below, displaying absence of CDX2 peaks in the *ST6GALNAC1* and *MUC1* respective loci. The former locus has previously been suggested to be regulated by CDX2[34], whereas the latter locus has not been linked to CDX2 regulation.



Supplementary Figure 7

Lineage and stemness gene expression profiles for LS-CDX2 PrIITE cells. RPKMs for selected epithelial marker genes and others are shown for LS174Twt (dark blue), CDX2$^{-/-}$2XindCDX2 without induction (red), CDX2$^{-/-}$2XindCDX2 with 40ng/ml Dox induction (green), CDX2$^{-/-}$2XindCDX2 with 4ng/ml Dox induction (purple) and CDX2$^{-/-}$1XindCDX2 with 4ng/ml Dox induction (light blue). RNA-seq expression patterns of intestinal and gastric specific marker genes in both LS174Twt and PrIITE cells, with or without induction were highly similar. Clearly the cells used and generated in this study displayed a mixed goblet, epithelial and stemcell expression profile, but notably lacked complete expression of the SI and ALPI enterocyte markers. Of notice, the gastric specific marker gene MUC6 was also found to be expressed in LS174Twt cells, in contrast *SOX2* was only found to be expressed in the CDX2$^{-/-}$ PrIITE cells generated.

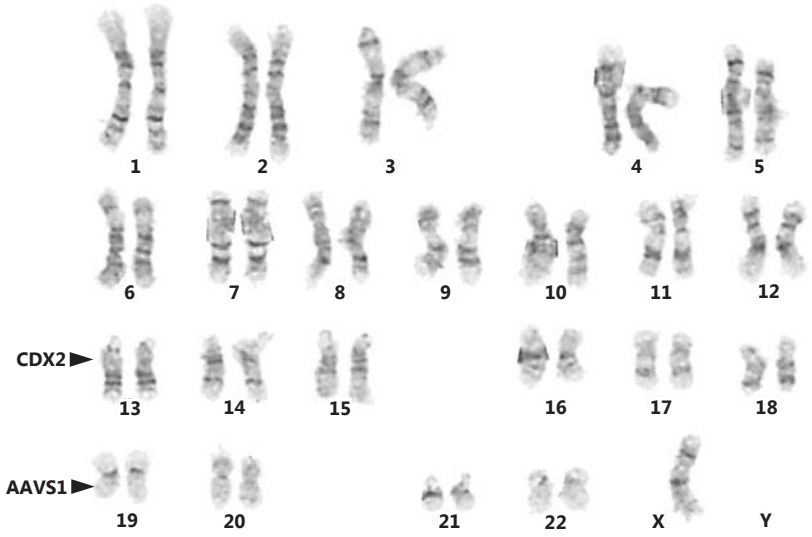
R Pinto et al., Supplementary Fig.8



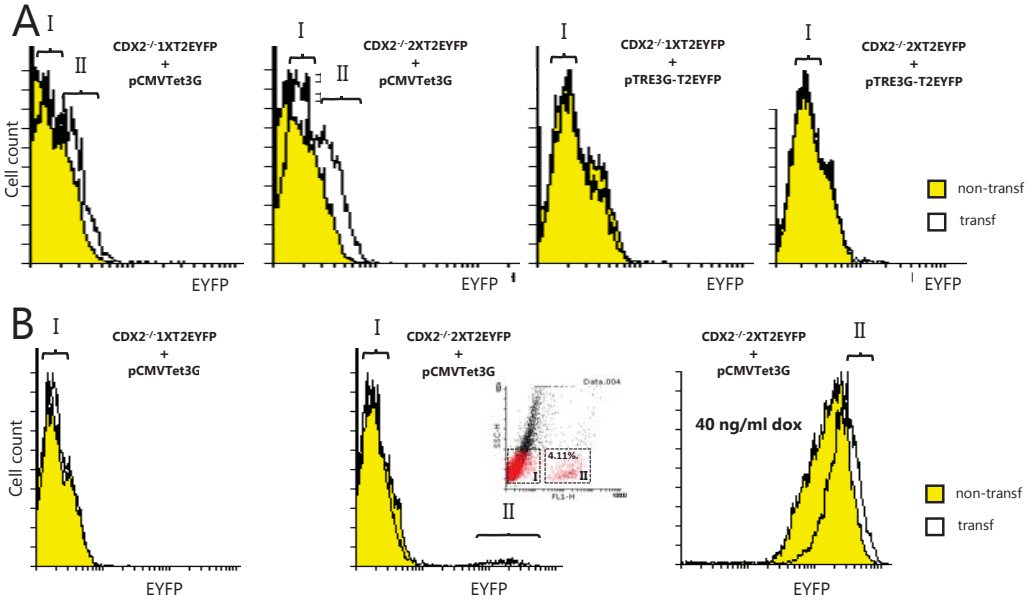
Supplementary Figure 8

Low and medium expressed LS174T genes regulated by CDX2. Targets genes downstream of CDX2 expressed at 50>RKPM>1 were analyzed based on the RNA-seq and ChIPSeq criterions 4X reduced and 2X increased RPKM in the respective uninduced and induced state and with CDX2 ChIPSeq peak in locus. The reversibility profiles for the genes in the respective PrIITE cells is similar as for the genes profiles displayed for highly expressed genes (RKPMs>50) shown in **Figure 7**, showing that 2XindCDX2 cells responded better than 1XindCDX2 cells. List of CDX2 regulated genes with 1>RKPMs >0 is given **Supplementary Table III. A.** Profiles for genes expressed with RPKMs between 10 and 50. **B.** Profiles for genes expressed with RPKMs between 10 and 1. **C.** Global Dox induction reversibility of the PrIITE cells relative to LS174Twt cells for low and medium level expressed genes (1<RPKM<50). Data analysis conditions as described in **Fig.7**, RNA-seq data (RPKM) from wildtype LS174T was plotted against RPKM values from uninduced CDX2$^{-/-}$2XindCDX2 cells without CDX2 expression (blue dotted line, 2XindCDX 0ng/ml DOX) and under different conditions of CDX2/Dox induction (ng/ml) (green dotted line, 2XindCDX2 4ng/ml DOX; purple dotted line, 1XindCDX2 4ng/ml DOX; red dotted line, 2XindCDX2 40ng/ml DOX). The regression line for each data set is plotted. The dark blue solid line represents the expression levels in LS174Twt cells. Representative examples for genes displaying near full reversibility are indicated with arrow heads. The data shows, that the induction profiles for low and medium expressed genes are similar to the global profiles shown in **Fig.7**.

R Pinto et al. Supplementary Fig.9

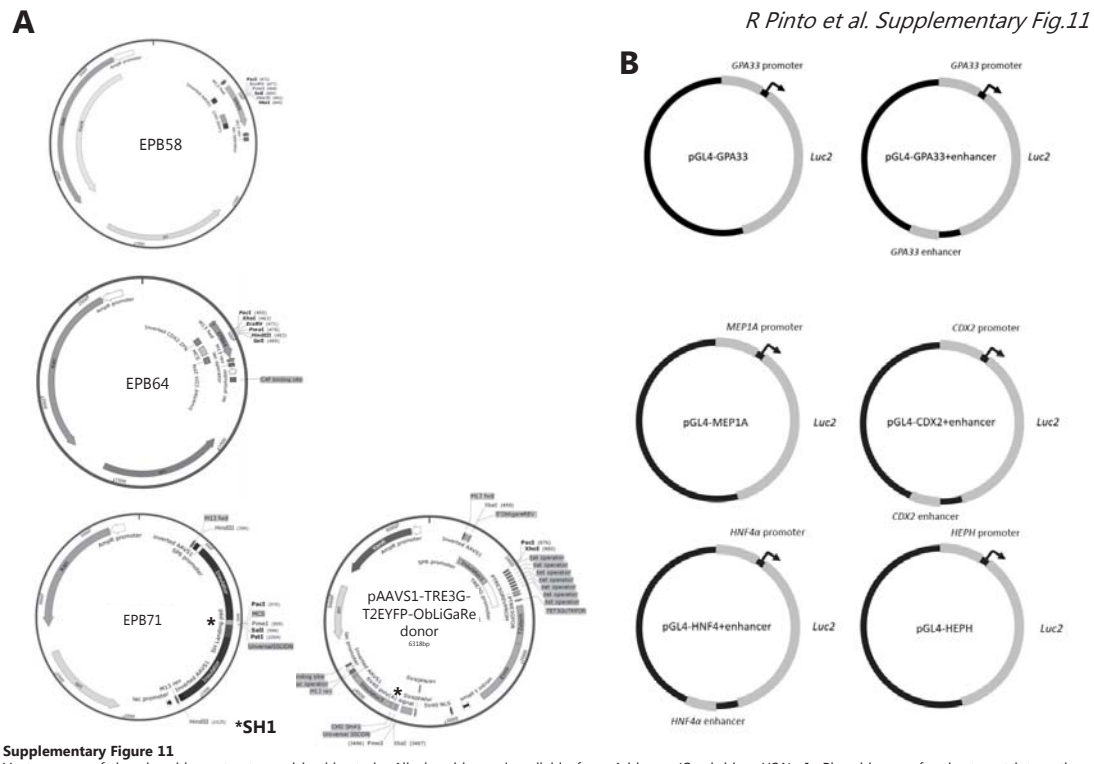


Supplemantary Figure 9
Karyogram of LS174T. The LS174T cell line used in this study was karyotyped using standard Giemsa-banding and karyotyping methods with slight modifications. Cells were treated with colcemid (0.004% w/v) for 1.5 hrs and trypsinized for 3-5 minutes before harvesting. All 25 cells analyzed from both the wild type and the modified cell line showed a 45,X karyotype with occasional 16q abnormality which was in agreement with the karyotype of the published LS174T cell line originating from a female adenocarcinoma (LP Rutzky et al., Cancer Research, 1980) displaying a near normal karyotype. The loci targeted in this study are indicated on the relevant chromosomes.



Supplementary Figure 10

Effect of increased expression and stable random integration on "leaky expression" of Tet3G inducible elements. **A.** LS174T PrIITE cells with defined precise targeted copies of TET3G and pTRE3G T2EYFP used as proof of principle, were transfected with either pCMVTet3G (left two panels) or pTRE3G-T2EYFP (right two panels) and the transiently transfected cell pools, 2 days post transfection, were analyzed by FACS for detection of potential "leaky" T2EYFP expression in the absence of Dox induction. Overlays of FACS analysis of non-transfected and transfected cells clearly showed, that transient overexpression of pCMVTet3G transactivator gave rise to two distinct population of cells; (I) a non fluorescent population identical to non transfected cells and (II) a population with increased leaky expression of T2EYFP distinct from the non-transfected cells. Transient over expression of the inducible pTRE3G-T2EYFP plasmid did not seem to give rise to leaky T2EYFP expression. **B.** Stable cell pools of the pCMVTet3G transfected cells analyzed in panel A, were obtained after up to 45 days G418 selection, followed by FACS analysis. Left panel shows that CDX2-/-1XT2EYFP cells do not display leaky expression as a consequence of increased expression of randomly integrated pCMVTE3G expression. Middle panel clearly shows that a 4.11% sub-population of cells (see side scatter insert for clarity) maintain substantial leaky T2EYFP expression in CDX2-/-2XT2EYFP cells at levels similar to 24h 40ng/ml Dox induction levels (right panel). Note, that the induction potential is increased in a stable randomly integrated pCMVTET3G sub-population of cells II). We speculated that these cells represent the cells with leaky expression in the non-induced state.



Supplementary Figure 11

Vector maps of the plasmid constructs used in this study. All plasmids used available from Addgene (Cambridge, USA). **A.** Plasmid maps for the target integration constructs used for establishing the various PrIITE cells. EPB58 (Addgene ID#90016) was used for insertion of TRE3G codon optimized CDX2 establishing *pAAVS1-TRE3G-CDX2opt ObLiGaRe-donor vector*. EPB64 (Addgene ID#90017) represents was used for inserting Tet3G transactivator into donor vector targeting the CDX2 exon 1 locus. EPB71 (Addgene ID#90018) used for establishing the T2EYFP AAVS1 donor integration vector pAAVS1-TRE3G-T2EYFP ObLiGaRe-donor vector shown. **B.** Vector maps for reporter plasmids used for promoter analysis of the various enhancer elements tested *in vitro*. Position of the CHO Safe Harbor #1 (SH1) "landing pad" is marked by an asterix.

Supplementary information:

Supplementary Table I. CDX2 as an activator*

Gene	LS174Twt CDX2$^{+/+}$ (rpkm)	CDX2$^{-/-}$2XindCDX2 0,00µg/ml dox	CDX2$^{-/-}$2XindCDX2 0,04µg/ml dox	CDX2$^{-/-}$2XindCDX2 0,004µg/ml dox	CDX2$^{+/-}$1XindCDX2 0,004µg/ml dox	>4 fold down	>2 fold up
TFF3	789,43	59,77	40,37	125,21	124,79	13,21	2,09
GPA33°	317,18	21,29	44,15	109,84	95,37	14,90	5,16
MUC2°	249,49	1,15	1,58	22,00	8,52	217,63	19,19
ETV4	164,05	22,13	45,55	89,86	64,74	7,41	4,06
ETHE1	110,69	25,75	40,56	71,57	65,55	4,30	2,78
LDLR	78,65	12,33	27,87	45,34	20,50	6,38	3,68
CDC42EP5	29,63	4,84	2,03	10,56	14,73	6,13	2,18
TMPRSS2	21,46	4,91	14,26	25,05	14,64	4,37	5,10
DFNB31	17,79	0,97	8,44	7,71	2,47	18,42	7,98
IRS2	11,67	2,32	5,38	19,35	10,03	5,02	8,32
TRIOBP	8,60	1,41	2,42	3,34	2,54	6,10	2,37
CALCA	7,09	0,69	4,00	5,80	0,79	10,22	8,35
ABCB1	6,57	0,22	3,89	2,11	1,83	30,38	9,78
CORO2A	6,20	1,08	5,23	5,72	2,61	5,77	5,31
PTK2B°	6,01	1,14	15,76	4,04	2,21	5,29	3,56
SLC6A20	5,53	0,59	2,67	8,73	6,97	9,42	14,86
SHH°	4,97	0,84	0,62	2,06	1,43	5,94	2,46
EGF	3,53	0,16	0,18	0,35	0,24	21,96	2,17

VPS37D	3,53	0,13	1,68	1,44	0,32	26,62	10,90
PDE9A°	2,66	0,17	0,33	0,80	0,64	15,69	4,71
THSD4	2,22	0,16	0,42	0,63	0,46	13,88	3,92
ATXN1	2,13	0,24	3,30	0,87	0,44	8,78	3,58
PTPRO	1,94	0,03	0,03	0,14	0,23	59,17	4,12
ARL14	1,58	0,36	6,91	3,87	1,74	4,42	10,78
B4GALNT3	1,48	0,26	1,16	1,58	0,75	5,79	6,15
LTBP1	1,46	0,10	0,59	0,42	0,28	13,88	3,96
KRT20	1,41	0,02	0,52	4,33	3,67	84,16	258,23
RBP3	1,20	0,04	0,01	0,26	0,10	27,75	6,06
CRB2	1,19	0,18	0,32	1,11	0,43	6,63	6,19
AOAH	1,15	0,01	0,63	1,15	0,22	133,41	132,89
CNBD1	1,01	0,01	0,02	0,19	0,15	83,96	15,82

* 4x↓in KO, 2x ↑+Dox (0.004µg/ml), CDX2 ChIP-seq peak in locus

° CDX2 targets previously identified in CaCo2 cells[1]

Supplementary TableII. CDX2 as a repressor*

Name	LS174Twt CDX2$^{+/+}$ (rpkm)	CDX2$^{-/-}$2XindCDX2 0,00µg/ml dox	CDX2$^{-/-}$2XindCDX2 0,04µg/ml dox	CDX2$^{-/-}$2XindCDX2 0,004µg/ml dox	CDX2$^{-/-}$1XindCDX2 0,004µg/ml dox
C4orf45	0	0,042776646	0	0	0,0242754
CD84	0	0,013292883	0,0094436	0,0075729	0,0037718
CFHR2	0	0,025168911	0	0	0,0285663
HOXA4	0	0,035210542	0,0166763	0	0,0199817
HOXC5	0	0,340791232	0,177545	0,0388293	0
IL17A	0	0,016544599	0	0	0
LIX1	0	0,022439014	0	0	0,0594251
MGAT4C	0	0,006539742	0,0123893	0	0
MORC1	0	0,008171203	0,00774	0	0,0092742
OR10V1	0	0,029516707	0	0	0
PTPRC	0	0,003615424	0,0205479	0	0,0041034
RPS6KA6	0	0,063854136	0,0302424	0	0,0362367
SLAMF6	0	0,302411884	0	0,0893313	0,3305203
TACR3	0	0,011852181	0,0112268	0,0067521	0
TEKT2	0	0,031098492	0,0441863	0	0

*Genes not expressed in LS174Twt, 4x ↑ in KO, 2x ↓+DOX (0.004µg/ml), CDX2 ChIP seq peak in locus

Supplementary Table III. CDX2 as an activator*

Name	LS174Twt CDX2$^{+/+}$ (rpkm)	CDX2$^{-/-}$2XindCDX2 0,00μg/ml dox	CDX2$^{-/-}$2XindCDX2 0,04μg/ml dox	CDX2$^{-/-}$2XindCDX2 0,004μg/ml dox	CDX2$^{-/-}$1XindCDX2 0,004μg dox
ANGPT1	0,7900189	5,5272512	3,9543323	4,6432396	1,9913809
C1orf21	0,7876576	3,2099251	1,9464836	3,0472531	3,4268473
AMACR	0,6386579	5,4593419	0,9981764	2,0324455	4,5895585
C9orf156	0,5126537	2,5889785	2,1962671	2,0630245	2,2875263
CABLES1	0,4026463	3,0876587	4,4746709	5,255985	3,3103341

*Low expressed genes 0<RPKM<1, 4x↓ in KO, 2x ↑+Dox induction (0,004μg/ml), CDX2 ChIP-seq peak in locus

Supplementary Table IV

	Sequence (5'->3')
Junction PCR	*CDX2* locus: CF: GGG CTC TCT GCT TGT CAC CTA CCA GG CR: CAC GCC CAT TGA TGT ACT GCC *AAVS1* locus: AF1: GCC CTC TAA CGC TGC CGT CTC AR2: TAT AGG CGC CCA CCG TAC ACG CC AF2: CACACCTCCCCCTGAACCTGA AR2: CGT AAG CAA ACC TTA GAG GTT CTG G
IDAA analysis	*CDX2* locus : CDX2S: AGC TGA CCG GCA GCA AAA TTG GAC GTG AGC ATG TAC CCT AGC TC CDX2AS: CTG CGC GCT GTC CAA GTT CGC TG *AAVS1* locus: AAVS1S: AGC TGA CCG GCA GCA AAA TTG CCT TAC CTC TCT AGT CTG TGC TAG AR2: CGT AAG CAA ACC TTA GAG GTT CTG G FAMF: FAM-AGC TGA CCG GCA GCA AAA TTG

3 Discussion

Maintaining the homeostasis in the colonic epithelium is important to avoid the development and exacerbation of IBD and CRC. The transcription factor network and the signalling pathways in the colon are crucial for maintaining the homeostasis. CDX2 is a main component of the transcription factor network. CDX2 interacts with both the Wnt and Hippo pathways to regulate target genes important for differentiation and homeostasis. Even though many of its target genes have been identified, it is still unclear how CDX2 affects colonic homeostasis and what its role is in diseases like IBD and CRC. Several studies have demonstrated that dysregulation of CDX2 is a factor in the development and severity of CRC and also to some extent in IBD (Calon et al., 2007; Coskun, 2014; Testa et al., 2018; Toth et al., 2018). The action of CDX2 is determined by its transcriptional regulation of genes, so knowing the target genes of CDX2 and how they are regulated improves the understanding of CDX2s role.

The starting point of this thesis work was bioinformatic analyses of CDX2 chromatin immunoprecipitated DNA from the colon epithelial cell lines Caco-2 and LS174T, which revealed gene regulatory elements within the *YAP1* and *IL33* genes with potential CDX2 binding sites. The roles of YAP1 and IL-33 in cell-proliferation and immune reactions are well studied, but their role in the colonic homeostasis and in IBD and CRC are still debated. Both genes are widely expressed in many tissues, but their regulation by an intestinal-specific transcription factor such as CDX2 indicate that they also could have intestinal-specific roles.

As a main downstream effector of the Hippo pathway, YAP1 modulates cell proliferation and organ size through its transcriptional co-activator effect and by its direct interaction with proteins of other signalling pathways in the colon. YAP1 has so far been demonstrated to have mostly oncogenic but also some tumour suppressor characteristics (Zygulska et al., 2017). The causes of its oncogenic properties are yet to be determined, but since YAP1 is involved in remodelling of the intestinal barrier and limiting of cell proliferation, these are areas of obvious interest. Defects in the colonic

epithelial barrier function is a characteristic feature of IBD (Antoni, Nuding, Wehkamp, & Stange, 2014). Upon intestinal tissue damage, it was found that YAP1 inhibits the Wnt homeostatic program while activating gene proliferation and that this can be a driver of cancer initiation (Gregorieff, Liu, Inanlou, Khomchuk, & Wrana, 2015). The role of IL-33 during colonic tissue damage remains unclear, but several studies have suggested that it may have a protective role as a transcription factor, by suppressing inflammation through inhibiting NF-κB signalling and the production of pro-inflammatory cytokines like IL-6 (Ali et al., 2011; Shao et al., 2014). This is in contrast to the well-known role of IL-33, where an intracellular IL-33 pool constitutes a store of alarmins that can be released during damage or inflammation of the intestinal barrier (Martin & Martin, 2016). In paper II, we showed that the amount of IL-33 protein in the media of LS174T colonic epithelial cells induced to express varying amounts of CDX2, was undetectable. No released IL-33 protein was detected even during high CDX2 levels and correspondingly high *IL33* mRNA levels. This suggests that IL-33 could be transported to the nucleus to act as a transcription factor, instead of being gathered in vesicles for release. However, the protein level was only measured in media from a monoculture of sterile and undamaged LS174T colon epithelial cells. Thus, it is possible that IL-33 protein could be released if the cells were damaged, co-cultured with pro-inflammatory immune cells, or stimulated with a secondary danger signal as is required for release of some other IL-1 family cytokines.

In paper I, the *YAP1* gene was investigated as a candidate CDX2 target gene in the colon. Bioinformatic analysis of CDX2 chromatin immunoprecipitated DNA from Caco-2 colon cells indicated that CDX2 binds to the *YAP1* promoter and to an intragenic regulatory element. Several CDX2 binding sites were identified in the promoter and intragenic element, and CDX2 binding to the sites was validated using RT-PCR quantification of Caco-2 ChIP DNA and EMSAs. Furthermore, transfections with *YAP1*-promoter reporter plasmids revealed the intragenic element to be a powerful enhancer that was able to activate expression from the *YAP1* promoter. The analyses also

showed that the enhancer activity was affected by the mutation of the CDX2 binding sites and that HNF4A is a powerful activator of the *YAP1* promoter activity. Interestingly, the protein expression analysis using the genetically engineered CDX2 inducible LS174T$^{CDX2-/-}$ colon cell model developed in Paper III showed that overexpressing CDX2 increased the YAP1 protein level. As expected, it also showed that YAP1 expression was present, as it is in most tissues, without CDX2 induction.

These results clearly demonstrate that CDX2 is able to regulate expression of the *YAP1* promoter and that the intragenic enhancer is important for *YAP1* expression. It was also shown that the CDX2 and HNF4A binding sites were important for the enhancers ability to upregulate *YAP1* promoter expression in Caco-2 cells and that CDX2 overexpression upregulates YAP1 expression. Given the role of CDX2 in *YAP1* regulation, it can be hypothesized that it confers an intestinal- or colon specific regulation of *YAP1* expression. This specificity allows the *YAP1* expression pattern to be tailored to the needs of colonic cells.

The investigated *YAP1* enhancer is inactive in LS174T cells, yet the CDX2 protein induction leads to increased YAP1 level. This suggests that other regulatory elements on the *YAP1* gene might be responsible for this increase in expression caused by CDX2 induction. Indeed, two additional potential regulatory elements were detected in the ChIP-Seq data, one in intron 1 and one in intron 4 (Paper I, Figure 1). It is plausible that these elements contribute to the effect of CDX2 in regulating YAP1 expression in colon cells.

Paper III describes the development of an LS174T colonic epithelial cell line genetically modified to contain a Tet3G-inducible system for the induction of CDX2. This model provides a way to investigate the effector functions of essential genes like CDX2 and to assess the dose-dependent effect of gene expression in a non-leaky system. The great value of this model system is the ability to have a relatively precise control over CDX2 expression. This is a clear advantage over other known colon cell lines such as Caco-2. The initial experimental work described in this thesis was performed using

Caco-2 cells. The reason for this was partly due to their capacity to differentiate in vitro and partly because many well-established CDX2 target genes have been identified and characterized using this cell line. However, CDX2 is essential for Caco-2 cell survival, and attempts at CDX2 knockdown by siRNA or lentiviral delivered shRNAs have been problematic in our hands. However, how LS174T cells have become viable without CDX2 expression is still unknown, and this advantage of the LS174T cells should entail some caution about the conclusions reached about the effector functions of re-introduced CDX2.

Understanding how molecular mechanisms, like transcriptional regulation, are controlled in the intestines is of great importance. For this purpose model systems are a crucial tool, as they can be manipulated and outcomes measured under strict laboratory conditions. Both Caco-2 and LS174T used in this thesis are well-established colonic epithelial cell lines that have been used to investigate intestinal regulation and cancer development for many years. However, these models are derived from colon cancer cells and thus contain numerous genetic alterations that allow them to survive *in vitro*. Even though most of these mutations have been identified, it is near impossible to predict the exact effect of these changes to the complex signalling pathways and resulting homeostasis of the cells, compared to healthy intestinal cells. To alleviate this problem other model systems have been developed. Another model to investigate colonic regulation could be to employ the recently developed colonic organoid model. Colonic organoids are three-dimensional *in vitro* cell models that incorporate many of the physiologically relevant features of the *in vivo* intestinal tissue. They can be isolated, grown, and expanded from intestinal tissue or they can be induced from pluripotent stem cells (Crespo et al., 2017; Munera et al., 2017; Sato et al., 2011). One of the advantages of using organoids instead of colon cancer cell lines is the ability to investigate the interaction of the different types of epithelial cells found in the crypts. However, this heterogeneity might also decrease the experimental specificity. Another disadvantage is that organoids still require a substantial effort, both time-wise and

monetary, over conventional cell lines. This might change, as they become a more established and widespread model system.

4 Conclusion

This thesis investigated the intestinal transcription factor CDX2s regulation of *YAP1* and *IL33* in colonic epithelial cells.

The first specific aim was to investigate whether the expression of *YAP1* is regulated by CDX2 in colonic epithelial cells, and to describe the mechanism of regulation. In paper I, a novel enhancer of *YAP1* expression was identified in Caco-2 cells. CDX2 was found to bind to the promoter and enhancer of *YAP1* and be important, for its expression. This is the first time that a mechanism of transcriptional regulation of *YAP1* has been described in colonic epithelial cells.

The second specific aim was to investigate whether the expression of *IL33* is regulated by CDX2 in colonic epithelial cells, and to describe the mechanism of regulation. In paper II, an intronic enhancer of *IL33* gene expression in LS174T colon epithelial cells was identified. Furthermore, CDX2 was found to be necessary for intestinal *IL33* expression and was found to up-regulate *IL33* expression when overexpressed in LS174T colonic epithelial cells.

The third specific aim was to develop a cell model that can be used to investigate the effects of varying CDX2 levels on target gene expression in colonic epithelial cells. Paper III describes how the precise integration of inducible transcriptional elements (PrIITE) method can be used to develop a CDX2 inducible LS174T colonic epithelial cell line. The LS174T$^{CDX2-/-}$ cells were shown to lack endogenous CDX2 expression until induced with doxycycline, showing no leakiness. Doxycycline stimulation was shown to induce CDX2 expression in a dose-dependent manner up to levels much higher than endogenous expression.

The overall aim of this thesis was to add new information about CDX2s regulation of colonic genes and to hypothesize about the potential effect of this regulation on homeostasis and diseases of the colon. The results presented in this thesis provides the first evidence that the intestinal specific transcription factor CDX2 regulates the expression of *YAP1* and *IL33* in colonic epithelial cells, thus adding novel

information about the role of CDX2 in the regulation of colonic genes. Together, the results discussed in this thesis brings us closer to understanding the role of CDX2, IL-33, and YAP1 in colon homeostasis and diseases of the colon.

5 Future perspectives

Most of the experiments included in the papers of this thesis were performed using LS174T and Caco-2 colon cancer epithelial cell lines. In paper II, it was shown that *IL33* mRNA level was significantly increased by induction of CDX2 in LS174T$^{CDX2-/-}$ cells and that no detectable level of IL-33 protein was released to the media. Detecting intracellular levels of accumulating IL-33 protein by either immunoblotting or immunostaining would support the hypothesis of IL-33 acting as a transcription factor in the homeostatic regulation of the intestinal epithelium. Furthermore, it could be interesting to test whether the colon epithelial cells release IL-33 in response to cell damage, co-culturing with immune cells, or direct stimulation with cytokines in the media.

The results from paper I demonstrated that the *YAP1* enhancer is active in Caco-2 cells but has no effect in LS174T cells. The reason behind this apparent cell-specific activity is unclear and worth investigating. It would be interesting to investigate the enhancer activity in several other colonic and intestinal cell lines. Additionally, the results show CDX2-dependent up-regulation of YAP1 protein in LS174T cells where the enhancer seems inactive. It would be relevant to determine if the two additional potential regulatory elements that were detected in the CDX2 ChIP-Seq data of paper I in intron 1 and 4 of *YAP1* could account for this up-regulation. This could be done by cloning the elements into promoter-reporter constructs and testing their effect on the *YAP1* promoter activity in CDX2 inducible LS174T$^{CDX2-/-}$ cells. Furthermore, it could be interesting to test the hypothesis that the YAP1 upregulation is caused by CDX2 regulating other transcription factors important for the activity of the enhancer. This could be accomplished by performing, for example, RNA-Seq or Chip-Seq on CDX2 induced and non-induced LS174T $^{CDX2-/-}$ cells and assessing differences in known transcription factors that are important for the colon homeostasis, like the HNF or GATA families.

6 References

Ali, S., Mohs, A., Thomas, M., Klare, J., Ross, R., Schmitz, M. L., & Martin, M. U. (2011). The dual function cytokine IL-33 interacts with the transcription factor NF-kappaB to dampen NF-kappaB-stimulated gene transcription. *J Immunol, 187*(4), 1609-1616. doi:10.4049/jimmunol.1003080

Antoni, L., Nuding, S., Wehkamp, J., & Stange, E. F. (2014). Intestinal barrier in inflammatory bowel disease. *World J Gastroenterol, 20*(5), 1165-1179. doi:10.3748/wjg.v20.i5.1165

Aoki, K., Kakizaki, F., Sakashita, H., Manabe, T., Aoki, M., & Taketo, M. M. (2011). Suppression of colonic polyposis by homeoprotein CDX2 through its nontranscriptional function that stabilizes p27Kip1. *Cancer Res, 71*(2), 593-602. doi:10.1158/0008-5472.CAN-10-2842

Aoki, K., Tamai, Y., Horiike, S., Oshima, M., & Taketo, M. M. (2003). Colonic polyposis caused by mTOR-mediated chromosomal instability in Apc+/Delta716 Cdx2+/- compound mutant mice. *Nat Genet, 35*(4), 323-330. doi:10.1038/ng1265

Attisano, L., & Wrana, J. L. (2013). Signal integration in TGF-β, WNT, and Hippo pathways. *F1000prime reports, 5*(June), 17-17. doi:10.12703/P5-17

Baekkevold, E. S., Roussigne, M., Yamanaka, T., Johansen, F. E., Jahnsen, F. L., Amalric, F., . . . Girard, J. P. (2003). Molecular characterization of NF-HEV, a nuclear factor preferentially expressed in human high endothelial venules. *Am J Pathol, 163*(1), 69-79. doi:10.1016/S0002-9440(10)63631-0

Bakaris, S., Cetinkaya, A., Ezberci, F., & Ekerbicer, H. (2008). Expression of homeodomain protein CDX2 in colorectal adenoma and adenocarcinoma. *Histology and histopathology, 23*(9), 1043-1047.

Barker, N., & Clevers, H. (2010). Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. *Gastroenterology, 138*(5), 1681-1696. doi:10.1053/j.gastro.2010.03.002

Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., . . . Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature, 449*(7165), 1003-1007. doi:10.1038/nature06196

Barros, R., da Costa, L. T., Pinto-de-Sousa, J., Duluc, I., Freund, J. N., David, L., & Almeida, R. (2011). CDX2 autoregulation in human intestinal metaplasia of the stomach: impact on the stability of the phenotype. *Gut, 60*(3), 290-298. doi:10.1136/gut.2010.222323

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell, 129*(4), 823-837. doi:10.1016/j.cell.2007.05.009

Basu, N., Saha, S., Khan, I., Ramachandra, S. G., & Visweswariah, S. S. (2014). Intestinal cell proliferation and senescence are regulated by receptor guanylyl

cyclase C and p21. *J Biol Chem, 289*(1), 581-593. doi:10.1074/jbc.M113.511311

Benahmed, F., Gross, I., Gaunt, S. J., Beck, F., Jehan, F., Domon-Dell, C., . . . Duluc, I. (2008). Multiple regulatory regions control the complex expression pattern of the mouse Cdx2 homeobox gene. *Gastroenterology, 135*(4), 1238-1247, 1247.e1231-1233. doi:10.1053/j.gastro.2008.06.045

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., . . . Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell, 120*(2), 169-181. doi:10.1016/j.cell.2005.01.001

Bhat, A. A., Sharma, A., Pope, J., Krishnan, M., Washington, M. K., Singh, A. B., & Dhawan, P. (2012). Caudal homeobox protein Cdx-2 cooperates with Wnt pathway to regulate claudin-1 expression in colon cancer cells. *PloS one, 7*(6), e37174-e37174. doi:10.1371/journal.pone.0037174

Biswas, S., Davis, H., Irshad, S., Sandberg, T., Worthley, D., & Leedham, S. (2015). Microenvironmental control of stem cell fate in intestinal homeostasis and disease. *J Pathol, 237*(2), 135-145. doi:10.1002/path.4563

Bonhomme, C., Calon, A., Martin, E., Robine, S., Neuville, A., Kedinger, M., . . . Freund, J. N. (2008). Cdx1, a dispensable homeobox gene for gut development with limited effect in intestinal cancer. *Oncogene, 27*(32), 4497-4502. doi:10.1038/onc.2008.78

Bonhomme, C., Duluc, I., Martin, E., Chawengsaksophak, K., Chenard, M. P., Kedinger, M., . . . Domon-Dell, C. (2003). The Cdx2 homeobox gene has a tumour suppressor function in the distal colon in addition to a homeotic role during gut development. *Gut, 52*(10), 1465-1471. doi:10.1136/gut.52.10.1465

Boulanger, J., Vezina, A., Mongrain, S., Boudreau, F., Perreault, N., Auclair, B. A., . . . Rivard, N. (2005). Cdk2-dependent phosphorylation of homeobox transcription factor CDX2 regulates its nuclear translocation and proteasome-mediated degradation in human intestinal epithelial cells. *J Biol Chem, 280*(18), 18095-18107. doi:10.1074/jbc.M502184200

Boyd, M., Bressendorff, S., Møller, J., Olsen, J., & Troelsen, J. T. (2009). Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC gastroenterology, 9*, 68-68. doi:10.1186/1471-230X-9-68

Boyd, M., Hansen, M., Jensen, T. G. K., Perearnau, A., Olsen, A. K., Bram, L. L., . . . Troelsen, J. T. (2010). Genome-wide analysis of CDX2 binding in intestinal epithelial cells (Caco-2). *The Journal of biological chemistry, 285*(33), 25115-25125. doi:10.1074/jbc.M109.089516

Braun, H., Afonina, I. S., Mueller, C., & Beyaert, R. (2018). Dichotomous function of IL-33 in health and disease: From biology to clinical implications. *Biochem Pharmacol, 148*, 238-252. doi:10.1016/j.bcp.2018.01.010

Brooks, A. K., & Gaj, T. (2018). Innovations in CRISPR technology. *Curr Opin Biotechnol, 52*, 95-101. doi:10.1016/j.copbio.2018.03.007

Buczacki, S. J., Zecchini, H. I., Nicholson, A. M., Russell, R., Vermeulen, L., Kemp, R., & Winton, D. J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature, 495*(7439), 65-69. doi:10.1038/nature11965

Calon, A., Gross, I., Lhermitte, B., Martin, E., Beck, F., Duclos, B., . . . Freund, J. N. (2007). Different effects of the Cdx1 and Cdx2 homeobox genes in a murine model of intestinal inflammation. *Gut, 56*(12), 1688-1695. doi:10.1136/gut.2007.125542

Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet, 6*(6), 507-512. doi:10.1038/nrg1619

Carriere, V., Roussel, L., Ortega, N., Lacorre, D. A., Americh, L., Aguilar, L., . . . Girard, J. P. (2007). IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo. *Proc Natl Acad Sci U S A, 104*(1), 282-287. doi:10.1073/pnas.0606854104

Cattin, A. L., Le Beyec, J., Barreau, F., Saint-Just, S., Houllier, A., Gonzalez, F. J., . . . Ribeiro, A. (2009). Hepatocyte nuclear factor 4alpha, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. *Mol Cell Biol, 29*(23), 6294-6308. doi:10.1128/MCB.00939-09

Cayrol, C., & Girard, J. P. (2014). IL-33: an alarmin cytokine with crucial roles in innate immunity, inflammation and allergy. *Curr Opin Immunol, 31*, 31-37. doi:10.1016/j.coi.2014.09.004

Chatterjee, S., & Ahituv, N. (2017). Gene Regulatory Elements, Major Drivers of Human Disease. *Annu Rev Genomics Hum Genet, 18*, 45-63. doi:10.1146/annurev-genom-091416-035537

Chawengsaksophak, K., James, R., Hammond, V. E., Köntgen, F., & Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature, 386*(6620), 84-87. doi:10.1038/386084a0

Chen, H. I., Einbond, A., Kwak, S. J., Linn, H., Koepf, E., Peterson, S., . . . Sudol, M. (1997). Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J Biol Chem, 272*(27), 17070-17077.

Choi, B. J., Kim, C. J., Cho, Y. G., Song, J. H., Kim, S. Y., Nam, S. W., . . . Park, W. S. (2006). Altered expression of CDX2 in colorectal cancers. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica, 114*(1), 50-54. doi:10.1111/j.1600-0463.2006.apm_292.x

Chou, S. T., Leng, Q., & Mixson, A. J. (2012). Zinc Finger Nucleases: Tailor-made for Gene Therapy. *Drugs Future, 37*(3), 183-196.

Clevers, H., & Nusse, R. (2012). Wnt/beta-catenin signaling and disease. *Cell, 149*(6), 1192-1205. doi:10.1016/j.cell.2012.05.012

Clevers, H. C., & Bevins, C. L. (2013). Paneth cells: maestros of the small intestinal crypts. *Annu Rev Physiol, 75*, 289-311. doi:10.1146/annurev-physiol-030212-183744

REFERENCES

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.), 339*(6121), 819-823. doi:10.1126/science.1231143

Coskun, M. (2014). The role of CDX2 in inflammatory bowel disease. *Dan Med J, 61*(3), B4820.

Coskun, M., Boyd, M., Olsen, J., & Troelsen, J. T. (2010). Control of intestinal promoter activity of the cellular migratory regulator gene ELMO3 by CDX2 and SP1. *Journal of cellular biochemistry, 109*(6), 1118-1128. doi:10.1002/jcb.22490

Coskun, M., Olsen, A. K., Bzorek, M., Holck, S., Engel, U. H., Nielsen, O. H., & Troelsen, J. T. (2014). Involvement of CDX2 in the cross talk between TNF- and Wnt signaling pathway in the colon cancer cell line Caco-2. *Carcinogenesis, 35*(5), 1185-1192. doi:10.1093/carcin/bgu037

Coskun, M., Olsen, A. K., Holm, T. L., Kvist, P. H., Nielsen, O. H., Riis, L. B., . . . Troelsen, J. T. (2012). TNF-α-induced down-regulation of CDX2 suppresses MEP1A expression in colitis. *Biochimica et biophysica acta, 1822*(6), 843-851. doi:10.1016/j.bbadis.2012.01.012

Coskun, M., Soendergaard, C., Joergensen, S., Dahlgaard, K., Riis, L. B., Nielsen, O. H., . . . Troelsen, J. T. (2017). Regulation of Laminin gamma2 Expression by CDX2 in Colonic Epithelial Cells Is Impaired During Active Inflammation. *J Cell Biochem, 118*(2), 298-307. doi:10.1002/jcb.25636

Crespo, M., Vilar, E., Tsai, S. Y., Chang, K., Amin, S., Srinivasan, T., . . . Chen, S. (2017). Colonic organoids derived from human induced pluripotent stem cells for modeling colorectal cancer and drug testing. *Nat Med, 23*(7), 878-884. doi:10.1038/nm.4355

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., . . . Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A, 107*(50), 21931-21936. doi:10.1073/pnas.1016071107

Cristina, M. L., Lehy, T., Zeitoun, P., & Dufougeray, F. (1978). Fine structural classification and comparative distribution of endocrine cells in normal human large intestine. *Gastroenterology, 75*(1), 20-28.

Dalerba, P., Sahoo, D., Paik, S., Guo, X., Yothers, G., Song, N., . . . Clarke, M. F. (2016). CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med, 374*(3), 211-222. doi:10.1056/NEJMoa1506597

Di Guglielmo, M. D., Park, J., Schulz, S., & Waldman, S. A. (2001). Nucleotide requirements for CDX2 binding to the cis promoter element mediating intestine-specific expression of guanylyl cyclase C. *FEBS Lett, 507*(2), 128-132.

Drewes, T., Senkel, S., Holewa, B., & Ryffel, G. U. (1996). Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Molecular and cellular biology, 16*(3), 925-931.

REFERENCES

Ekbom, A., Helmick, C., Zack, M., & Adami, H. O. (1990). Ulcerative colitis and colorectal cancer. A population-based study. *N Engl J Med, 323*(18), 1228-1233. doi:10.1056/NEJM199011013231802

Engholm G, F. J., Christensen N, Hansen HL, Hertzum-Larsen R, Johannesen TB, Kejs AMT, Khan S, Ólafsdóttir E, Petersen T, Schmidt LKH, Virtanen A and Storm HH. (2018). Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.1 (28.06.2018). Association of the Nordic Cancer Registries. Danish Cancer Society. Available from http://www.ancr.nu, accessed on 01/08/2018. .

Epinat, J. C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., . . . Lacroix, E. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res, 31*(11), 2952-2962.

Fedotova, A. A., Bonchuk, A. N., Mogila, V. A., & Georgiev, P. G. (2017). C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta Naturae, 9*(2), 47-58.

Frietze, S., & Farnham, P. J. (2011). Transcription factor effector domains. *Subcell Biochem, 52*, 261-277. doi:10.1007/978-90-481-9069-0_12

Gao, N., White, P., & Kaestner, K. H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev Cell, 16*(4), 588-599. doi:10.1016/j.devcel.2009.02.010

Garrison, W. D., Battle, M. A., Yang, C., Kaestner, K. H., Sladek, F. M., & Duncan, S. A. (2006). Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. *Gastroenterology, 130*(4), 1207-1220. doi:10.1053/j.gastro.2006.01.003

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America, 109*(39), E2579-2586. doi:10.1073/pnas.1208507109

Gautier-Stein, A., Zitoun, C., Lalli, E., Mithieux, G., & Rajas, F. (2006). Transcriptional regulation of the glucose-6-phosphatase gene by cAMP/vasoactive intestinal peptide in the intestine. Role of HNF4alpha, CREM, HNF1alpha, and C/EBPalpha. *The Journal of biological chemistry, 281*(42), 31268-31278. doi:10.1074/jbc.M603258200

Gilbert, L. a., Larson, M. H., Morsut, L., Liu, Z., Brar, G. a., Torres, S. E., . . . Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell, 154*(2), 442-451. doi:10.1016/j.cell.2013.06.044

Goth, R., & Rajewsky, M. F. (1972). Ethylation of nucleic acids by ethylnitrosourea-1-14 C in the fetal and adult rat. *Cancer Res, 32*(7), 1501-1505.

Gregorieff, A., Liu, Y., Inanlou, M. R., Khomchuk, Y., & Wrana, J. L. (2015). Yap-dependent reprogramming of Lgr5(+) stem cells drives intestinal regeneration and cancer. *Nature, 526*(7575), 715-718. doi:10.1038/nature15382

Grobeta, P., Doser, K., Falk, W., Obermeier, F., & Hofmann, C. (2012). IL-33 attenuates development and perpetuation of chronic intestinal inflammation. *Inflamm Bowel Dis, 18*(10), 1900-1909. doi:10.1002/ibd.22900

Gunawardene, A. R., Corfe, B. M., & Staton, C. A. (2011). Classification and functions of enteroendocrine cells of the lower gastrointestinal tract. *Int J Exp Pathol, 92*(4), 219-231. doi:10.1111/j.1365-2613.2011.00767.x

Guo, R.-J., Funakoshi, S., Lee, H. H., Kong, J., & Lynch, J. P. (2010). The intestine-specific transcription factor Cdx2 inhibits β-catenin/TCF transcriptional activity by disrupting the β-catenin–TCF protein complex. *Carcinogenesis, 31*(2), 159-166. doi:10.1093/carcin/bgp213

Haggar, F. A., & Boushey, R. P. (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg, 22*(4), 191-197. doi:10.1055/s-0029-1242458

Hill, M. J. (1997). Intestinal flora and endogenous vitamin synthesis. *Eur J Cancer Prev, 6 Suppl 1*, S43-45.

Hinkel, I., Duluc, I., Martin, E., Guenot, D., Freund, J. N., & Gross, I. (2012). Cdx2 controls expression of the protocadherin Mucdhl, an inhibitor of growth and beta-catenin activity in colon cancer cells. *Gastroenterology, 142*(4), 875-885 e873. doi:10.1053/j.gastro.2011.12.037

Hinoi, T., Gesina, G., Akyol, A., Kuick, R., Hanash, S., Giordano, T. J., . . . Fearon, E. R. (2005). CDX2-regulated expression of iron transport protein hephaestin in intestinal and colonic epithelium. *Gastroenterology, 128*(4), 946-961.

Hinoi, T., Lucas, P. C., Kuick, R., Hanash, S., Cho, K. R., & Fearon, E. R. (2002). CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. *Gastroenterology, 123*(5), 1565-1577.

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science, 327*(5962), 167-170. doi:10.1126/science.1179555

Houde, M., Laprise, P., Jean, D., Blais, M., Asselin, C., & Rivard, N. (2001). Intestinal epithelial cell differentiation involves activation of p38 mitogen-activated protein kinase that regulates the homeobox transcription factor CDX2. *J Biol Chem, 276*(24), 21885-21894. doi:10.1074/jbc.M100236200

Hryniuk, A., Grainger, S., Savory, J. G., & Lohnes, D. (2012). Cdx function is required for maintenance of intestinal identity in the adult. *Dev Biol, 363*(2), 426-437. doi:10.1016/j.ydbio.2012.01.010

Humphries, A., & Wright, N. a. (2008). Colonic crypt organization and tumorigenesis. *Nature reviews. Cancer, 8*(6), 415-424. doi:10.1038/nrc2392

Hussain, M. A., & Habener, J. F. (1999). Glucagon gene transcription activation mediated by synergistic interactions of pax-6 and cdx-2 with the p300 co-activator. *The Journal of biological chemistry, 274*(41), 28950-28957.

Imaeda, H., Andoh, A., Aomatsu, T., Uchiyama, K., Bamba, S., Tsujikawa, T., . . . Fujiyama, Y. (2011). Interleukin-33 suppresses Notch ligand expression and

prevents goblet cell depletion in dextran sulfate sodium-induced colitis. *Int J Mol Med, 28*(4), 573-578. doi:10.3892/ijmm.2011.718

Imajo, M., Miyatake, K., Iimura, A., Miyamoto, A., & Nishida, E. (2012). A molecular mechanism that links Hippo signalling to the inhibition of Wnt/β-catenin signalling. *The EMBO journal, 31*(5), 1109-1122. doi:10.1038/emboj.2011.487

James, R., & Kazenwadel, J. (1991). Homeobox gene expression in the intestinal epithelium of adult mice. *The Journal of biological chemistry, 266*(5), 3246-3251.

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science, 293*(5532), 1074-1080. doi:10.1126/science.1063127

Jess, T., Gamborg, M., Matzen, P., Munkholm, P., & Sorensen, T. I. (2005). Increased risk of intestinal cancer in Crohn's disease: a meta-analysis of population-based cohort studies. *Am J Gastroenterol, 100*(12), 2724-2729. doi:10.1111/j.1572-0241.2005.00287.x

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.), 337*(6096), 816-821. doi:10.1126/science.1225829

Johansson, M. E., Larsson, J. M., & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A, 108 Suppl 1*, 4659-4665. doi:10.1073/pnas.1006451107

Johansson, M. E., Thomsson, K. A., & Hansson, G. C. (2009). Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the Fcgbp protein. *J Proteome Res, 8*(7), 3549-3557. doi:10.1021/pr9002504

Kim, J. S. (2016). Genome editing comes of age. *Nat Protoc, 11*(9), 1573-1578. doi:10.1038/nprot.2016.104

Kim, S. P., Park, J. W., Lee, S. H., Lim, J. H., Jang, B. C., Lee, S. H., . . . Kwon, T. K. (2004). Homeodomain protein CDX2 regulates COX-2 expression in colorectal cancer. *Biochem Biophys Res Commun, 315*(1), 93-99. doi:10.1016/j.bbrc.2004.01.020

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A, 93*(3), 1156-1160.

Kim, Y. G., & Chandrasegaran, S. (1994). Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A, 91*(3), 883-887.

Kobori, A., Yagi, Y., Imaeda, H., Ban, H., Bamba, S., Tsujikawa, T., . . . Andoh, A. (2010). Interleukin-33 expression is specifically enhanced in inflamed mucosa of ulcerative colitis. *J Gastroenterol, 45*(10), 999-1007. doi:10.1007/s00535-010-0245-1

REFERENCES

Kosinski, C., Li, V. S., Chan, A. S., Zhang, J., Ho, C., Tsui, W. Y., . . . Chen, X. (2007). Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc Natl Acad Sci U S A, 104*(39), 15418-15423. doi:10.1073/pnas.0707210104

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harb Protoc, 2015*(11), 951-969. doi:10.1101/pdb.top084970

Li, Z., Zhao, B., Wang, P., Chen, F., Dong, Z., Yang, H., . . . Xu, Y. (2010). Structural insights into the YAP and TEAD complex. *Genes Dev, 24*(3), 235-240. doi:10.1101/gad.1865810

Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., . . . Guan, K. L. (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev, 24*(11), 1106-1118. doi:10.1101/gad.1903310

Liang, K., Zhou, G., Zhang, Q., Li, J., & Zhang, C. (2014). Expression of hippo pathway in colorectal cancer. *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association, 20*(3), 188-194. doi:10.4103/1319-3767.133025

Lingel, A., Weiss, T. M., Niebuhr, M., Pan, B., Appleton, B. A., Wiesmann, C., . . . Fairbrother, W. J. (2009). Structure of IL-33 and its interaction with the ST2 and IL-1RAcP receptors--insight into heterotrimeric IL-1 signaling complexes. *Structure, 17*(10), 1398-1410. doi:10.1016/j.str.2009.08.009

Lipkin, M., Bell, B., & Sherlock, P. (1963). Cell Proliferation Kinetics in the Gastrointestinal Tract of Man. I. Cell Renewal in Colon and Rectum. *J Clin Invest, 42*(6), 767-776. doi:10.1172/JCI104769

Liu, X., Zhu, L., Lu, X., Bian, H., Wu, X., Yang, W., & Qin, Q. (2014). IL-33/ST2 pathway contributes to metastasis of human colorectal cancer. *Biochem Biophys Res Commun, 453*(3), 486-492. doi:10.1016/j.bbrc.2014.09.106

M. Ervik, F. L., J. Ferlay, L. Mery, I. Soerjomataram, F. Bray. (2016). Cancer Today. Lyon, France: International Agency for Research on Cancer. Cancer Today. Available from: http://gco.iarc.fr/today, accessed [01/08/2018]. .

Ma, J. (2011). Transcriptional activators and activation mechanisms. *Protein Cell, 2*(11), 879-888. doi:10.1007/s13238-011-1101-7

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., & Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature Methods, 10*(10), 977-979. doi:10.1038/nmeth.2598

Martin, N. T., & Martin, M. U. (2016). Interleukin 33 is a guardian of barriers and a local alarmin. *Nat Immunol, 17*(2), 122-131. doi:10.1038/ni.3370

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., . . . Rebar, E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol, 29*(2), 143-148. doi:10.1038/nbt.1755

Mitchell, P. J., & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science, 245*(4916), 371-378.

Mitchelmore, C., Troelsen, J. T., Spodsberg, N., Sjöström, H., & Norén, O. (2000). Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. *The Biochemical journal, 346 Pt 2*, 529-535.

Mito, A., Nakano, Y., Saitoh, T., Gouraud, S. S. S., Yamaguchi, Y., Sato, T., . . . Kojima-Aikawa, K. (2018). Lectin ZG16p inhibits proliferation of human colorectal cancer cells via its carbohydrate-binding sites. *Glycobiology, 28*(1), 21-31. doi:10.1093/glycob/cwx088

Moussion, C., Ortega, N., & Girard, J. P. (2008). The IL-1-like cytokine IL-33 is constitutively expressed in the nucleus of endothelial cells and epithelial cells in vivo: a novel 'alarmin'? *PloS one, 3*(10), e3331. doi:10.1371/journal.pone.0003331

Munera, J. O., Sundaram, N., Rankin, S. A., Hill, D., Watson, C., Mahe, M., . . . Wells, J. M. (2017). Differentiation of Human Pluripotent Stem Cells into Colonic Organoids via Transient Activation of BMP Signaling. *Cell Stem Cell, 21*(1), 51-64 e56. doi:10.1016/j.stem.2017.05.020

Mutoh, H., Hayakawa, H., Sakamoto, H., & Sugano, K. (2007). Homeobox protein CDX2 reduces Cox-2 transcription by inactivating the DNA-binding capacity of nuclear factor-kappaB. *J Gastroenterol, 42*(9), 719-729. doi:10.1007/s00535-007-2088-y

Ng, S. C., Shi, H. Y., Hamidi, N., Underwood, F. E., Tang, W., Benchimol, E. I., . . . Kaplan, G. G. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet, 390*(10114), 2769-2778. doi:10.1016/S0140-6736(17)32448-0

Nguyen, L. T., Tretiakova, M. S., Silvis, M. R., Lucas, J., Klezovitch, O., Coleman, I., . . . Vasioukhin, V. (2015). ERG Activates the YAP1 Transcriptional Program and Induces the Development of Age-Related Prostate Tumors. *Cancer Cell, 27*(6), 797-808. doi:10.1016/j.ccell.2015.05.005

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., & Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell, 123*(5), 917-929. doi:10.1016/j.cell.2005.08.040

Noah, T. K., Donahue, B., & Shroyer, N. F. (2011). Intestinal development and differentiation. *Exp Cell Res, 317*(19), 2702-2710. doi:10.1016/j.yexcr.2011.09.006

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., & Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell, 133*(7), 1277-1289. doi:10.1016/j.cell.2008.05.023

Naar, A. M., Lemon, B. D., & Tjian, R. (2001). Transcriptional coactivator complexes. *Annu Rev Biochem, 70*, 475-501. doi:10.1146/annurev.biochem.70.1.475

Oboki, K., Ohno, T., Kajiwara, N., Arae, K., Morita, H., Ishii, A., . . . Nakae, S. (2010). IL-33 is a crucial amplifier of innate rather than acquired immunity. *Proc Natl Acad Sci U S A, 107*(43), 18581-18586. doi:10.1073/pnas.1003059107

Olsen, A. K., Boyd, M., Danielsen, E. T., & Troelsen, J. T. (2012). Current and emerging approaches to define intestinal epithelium-specific transcriptional networks. *American journal of physiology. Gastrointestinal and liver physiology, 302*(3), G277-286. doi:10.1152/ajpgi.00362.2011

Pan, D. (2010). The hippo signaling pathway in development and cancer. *Developmental Cell, 19*(4), 491-505. doi:10.1016/j.devcel.2010.09.011

Panne, D. (2008). The enhanceosome. *Curr Opin Struct Biol, 18*(2), 236-242. doi:10.1016/j.sbi.2007.12.002

Pelaseyed, T., Bergstrom, J. H., Gustafsson, J. K., Ermund, A., Birchenough, G. M., Schutte, A., . . . Hansson, G. C. (2014). The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev, 260*(1), 8-20. doi:10.1111/imr.12182

Porteus, M. H., & Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science, 300*(5620), 763. doi:10.1126/science.1078395

Pushparaj, P. N., Li, D., Komai-Koma, M., Guabiraba, R., Alexander, J., McSharry, C., & Xu, D. (2013). Interleukin-33 exacerbates acute colitis via interleukin-4 in mice. *Immunology, 140*(1), 70-77. doi:10.1111/imm.12111

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell, 152*(5), 1173-1183. doi:10.1016/j.cell.2013.02.022

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature, 470*(7333), 279-283. doi:10.1038/nature09692

Ralston, A., & Rossant, J. (2008). Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev Biol, 313*(2), 614-629. doi:10.1016/j.ydbio.2007.10.054

Reddy, M. A., Park, J. T., & Natarajan, R. (2012). Epigenetic modifications and diabetic nephropathy. *Kidney Res Clin Pract, 31*(3), 139-150. doi:10.1016/j.krcp.2012.07.004

Renouf, B., Soret, C., Saandi, T., Delalande, F., Martin, E., Vanier, M., . . . Domon-Dell, C. (2012). Cdx2 homeoprotein inhibits non-homologous end joining in colon cancer but not in leukemia cells. *Nucleic Acids Res, 40*(8), 3456-3469. doi:10.1093/nar/gkr1242

Rings, E. H. H. M., Boudreau, F., Taylor, J. K., Moffett, J., Suh, E. R., & Traber, P. G. (2001). Phosphorylation of the serine 60 residue within the Cdx2 activation domain mediates its transactivation capacity. *Gastroenterology, 121*(6), 1437-1450. doi:10.1053/gast.2001.29618

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., . . . Kent, W. J. (2013). ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research, 41*(Database issue), D56-63. doi:10.1093/nar/gks1172

Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol, 14*(12), 8096-8106.

Roussel, L., Erard, M., Cayrol, C., & Girard, J. P. (2008). Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A-H2B acidic pocket. *EMBO Rep, 9*(10), 1006-1012. doi:10.1038/embor.2008.145

Rudin, N., & Haber, J. E. (1988). Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. *Mol Cell Biol, 8*(9), 3918-3928.

Ryu, J.-H., Kim, S.-H., Lee, H.-Y., Bai, J. Y., Nam, Y.-D., Bae, J.-W., . . . Lee, W.-J. (2008). Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in Drosophila. *Science (New York, N.Y.), 319*(5864), 777-782. doi:10.1126/science.1149357

Said, H. M. (2013). Recent advances in transport of water-soluble vitamins in organs of the digestive system: a focus on the colon and the pancreas. *Am J Physiol Gastrointest Liver Physiol, 305*(9), G601-610. doi:10.1152/ajpgi.00231.2013

Saksouk, N., Simboeck, E., & Dejardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin, 8*, 3. doi:10.1186/1756-8935-8-3

Salari, K., Spulak, M. E., Cuff, J., Forster, A. D., Giacomini, C. P., Huang, S., . . . Pollack, J. R. (2012). CDX2 is an amplified lineage-survival oncogene in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America, 109*(46), E3196-E3205. doi:10.1073/pnas.1206004109

San Roman, A. K., Aronson, B. E., Krasinski, S. D., Shivdasani, R. a., & Verzi, M. P. (2015). Transcription Factors GATA4 and HNF4A Control Distinct Aspects of Intestinal Homeostasis in Conjunction with Transcription Factor CDX2. *Journal of Biological Chemistry, 290*(3), 1850-1860. doi:10.1074/jbc.M114.620211

Sancho, E., Batlle, E., & Clevers, H. (2004). Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol, 20*, 695-723. doi:10.1146/annurev.cellbio.20.010403.092805

Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S. I., Fumagalli, A., Lyubimova, A., . . . Clevers, H. (2016). Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proc Natl Acad Sci U S A, 113*(37), E5399-5407. doi:10.1073/pnas.1607327113

Sato, T., Stange, D. E., Ferrante, M., Vries, R. G., Van Es, J. H., Van den Brink, S., . . . Clevers, H. (2011). Long-term expansion of epithelial organoids from human

colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology, 141*(5), 1762-1772. doi:10.1053/j.gastro.2011.07.050

Sedhom, M. A., Pichery, M., Murdoch, J. R., Foligne, B., Ortega, N., Normand, S., . . . Ryffel, B. (2013). Neutralisation of the interleukin-33/ST2 pathway ameliorates experimental colitis through enhancement of mucosal healing in mice. *Gut, 62*(12), 1714-1723. doi:10.1136/gutjnl-2011-301785

Seidelin, J. B., Bjerrum, J. T., Coskun, M., Widjaya, B., Vainer, B., & Nielsen, O. H. (2010). IL-33 is upregulated in colonocytes of ulcerative colitis. *Immunol Lett, 128*(1), 80-85. doi:10.1016/j.imlet.2009.11.001

Seo, E., Basu-Roy, U., Gunaratne, P. H., Coarfa, C., Lim, D.-S. S., Basilico, C., & Mansukhani, A. (2013). SOX2 regulates YAP1 to maintain stemness and determine cell fate in the osteo-adipo lineage. *Cell Reports, 3*(6), 2075-2087. doi:10.1016/j.celrep.2013.05.029

Shao, D., Perros, F., Caramori, G., Meng, C., Dormuller, P., Chou, P. C., . . . Wort, S. J. (2014). Nuclear IL-33 regulates soluble ST2 receptor and IL-6 expression in primary human arterial endothelial cells and is decreased in idiopathic pulmonary arterial hypertension. *Biochem Biophys Res Commun, 451*(1), 8-14. doi:10.1016/j.bbrc.2014.06.111

Silberg, D. G., Swain, G. P., Suh, E. R., & Traber, P. G. (2000). Cdx1 and cdx2 expression during intestinal development. *Gastroenterology, 119*(4), 961-971.

Simmini, S., Bialecka, M., Huch, M., Kester, L., van de Wetering, M., Sato, T., . . . Deschamps, J. (2014). Transformation of intestinal stem cells into gastric stem cells on loss of transcription factor Cdx2. *Nature communications, 5*, 5728-5728. doi:10.1038/ncomms6728

Sjolund, K., Sanden, G., Hakanson, R., & Sundler, F. (1983). Endocrine cells in human intestine: an immunocytochemical study. *Gastroenterology, 85*(5), 1120-1130.

Smith, A. J. P., Deloukas, P., & Munroe, P. B. (2018). Emerging applications of genome-editing technology to examine functionality of GWAS-associated variants for complex traits. *Physiol Genomics, 50*(7), 510-522. doi:10.1152/physiolgenomics.00028.2018

Song, L., & Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc, 2010*(2), pdb prot5384. doi:10.1101/pdb.prot5384

Stein, C., Bardet, A. F., Roma, G., Bergling, S., Clay, I., Ruchti, A., . . . Bauer, A. (2015). YAP1 Exerts Its Transcriptional Control via TEAD-Mediated Activation of Enhancers. *PLoS Genet, 11*(8), e1005465. doi:10.1371/journal.pgen.1005465

Steinhardt, A. A., Gayyed, M. F., Klein, A. P., Dong, J., Maitra, A., Pan, D., . . . Anders, R. A. (2008). Expression of Yes-associated protein in common solid tumors. *Human pathology, 39*(11), 1582-1589. doi:10.1016/j.humpath.2008.04.012

Suh, E., & Traber, P. G. (1996). An intestine-specific homeobox gene regulates proliferation and differentiation. *Molecular and cellular biology, 16*(2), 619-625.

Talabot-Ayer, D., Lamacchia, C., Gabay, C., & Palmer, G. (2009). Interleukin-33 is biologically active independently of caspase-1 cleavage. *J Biol Chem, 284*(29), 19420-19426. doi:10.1074/jbc.M901744200

Testa, U., Pelosi, E., & Castelli, G. (2018). Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells. *Med Sci (Basel), 6*(2). doi:10.3390/medsci6020031

Toth, C., Sukosd, F., Valicsek, E., Herpel, E., Schirmacher, P., & Tiszlavicz, L. (2018). Loss of CDX2 gene expression is associated with DNA repair proteins and is a crucial member of the Wnt signaling pathway in liver metastasis of colorectal cancer. *Oncol Lett, 15*(3), 3586-3593. doi:10.3892/ol.2018.7756

Troelsen, J. T., Mitchelmore, C., Spodsberg, N., Jensen, A. M., Noren, O., & Sjostrom, H. (1997). Regulation of lactase-phlorizin hydrolase gene expression by the caudal-related homoeodomain protein Cdx-2. *Biochem J, 322 (Pt 3)*, 833-838.

Tschaharganeh, D. F., Chen, X., Latzko, P., Malz, M., Gaida, M. M., Felix, K., . . . Breuhahn, K. (2013). Yes-associated protein up-regulates Jagged-1 and activates the Notch pathway in human hepatocellular carcinoma. *Gastroenterology, 144*(7), 1530-1542.e1512. doi:10.1053/j.gastro.2013.02.009

Tsubouchi, S., & Leblond, C. P. (1979). Migration and turnover of entero-endocrine and caveolated cells in the epithelium of the descending colon, as shown by radioautography after continuous infusion of 3H-thymidine into mice. *Am J Anat, 156*(4), 431-451. doi:10.1002/aja.1001560403

Tupler, R., Perini, G., & Green, M. R. (2001). Expressing the human genome. *Nature, 409*(6822), 832-833. doi:10.1038/35057011

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science, 347*(6220), 1260419. doi:10.1126/science.1260419

Verzi, M. P., Shin, H., He, H. H., Sulahian, R., Meyer, C. A., Montgomery, R. K., . . . Shivdasani, R. A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Developmental Cell, 19*(5), 713-726. doi:10.1016/j.devcel.2010.10.006

Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W., & Povey, S. (2002). Guidelines for human gene nomenclature. *Genomics, 79*(4), 464-470. doi:10.1006/geno.2002.6748

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet, 10*(1), 57-63. doi:10.1038/nrg2484

Wasmer, M. H., & Krebs, P. (2016). The Role of IL-33-Dependent Inflammation in the Tumor Microenvironment. *Front Immunol, 7*, 682. doi:10.3389/fimmu.2016.00682

Wiedenheft, B., Sternberg, S. H., & Doudna, J. a. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature, 482*(7385), 331-338. doi:10.1038/nature10886

Wilson, L. O. W., O'Brien, A. R., & Bauer, D. C. (2018). The Current State and Future of CRISPR-Cas9 gRNA Design Tools. *Front Pharmacol, 9*, 749. doi:10.3389/fphar.2018.00749

Witek, M. E., Snook, A. E., Lin, J. E., Blomain, E. S., Xiang, B., Magee, M. S., & Waldman, S. A. (2014). A novel CDX2 isoform regulates alternative splicing. *PloS one, 9*(8), e104293. doi:10.1371/journal.pone.0104293

Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N., & van der Oost, J. (2018). Genome editing by natural and engineered CRISPR-associated nucleases. *Nat Chem Biol, 14*(7), 642-651. doi:10.1038/s41589-018-0080-x

Wunderlich, Z., & Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet, 25*(10), 434-440. doi:10.1016/j.tig.2009.08.003

Xu, F. (1999). Cell Type-specific Autoregulation of the Caudal-related Homeobox Gene Cdx-2/3. *Journal of Biological Chemistry, 274*(48), 34310-34316. doi:10.1074/jbc.274.48.34310

Yui, S., Azzolin, L., Maimets, M., Pedersen, M. T., Fordham, R. P., Hansen, S. L., . . . Jensen, K. B. (2018). YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell, 22*(1), 35-49 e37. doi:10.1016/j.stem.2017.11.001

Zhao, B., Li, L., & Guan, K. L. (2010). Hippo signaling at a glance. *J Cell Sci, 123*(Pt 23), 4001-4006. doi:10.1242/jcs.069070

Zhao, B., Li, L., Lei, Q., & Guan, K. L. (2010). The Hippo–YAP pathway in organ size control and tumorigenesis: An updated version. *Genes & development, 24*, 862-874. doi:10.1101/gad.1909210

Zhu, J., Yang, F., Sang, L., Zhai, J., Zhang, X., Yue, D., . . . Sun, X. (2015). IL-33 Aggravates DSS-Induced Acute Colitis in Mouse Colon Lamina Propria by Enhancing Th2 Cell Responses. *Mediators Inflamm, 2015*, 913041. doi:10.1155/2015/913041

Zygulska, A. L., Krzemieniecki, K., & Pierzchalski, P. (2017). Hippo pathway - brief overview of its relevance in cancer. *J Physiol Pharmacol, 68*(3), 311-335.