Study of RPC32α, subunit of the RNA polymerase III, in a tumor model

Wiebke Bretting

▶ **To cite this version:**

Wiebke Bretting. Study of RPC32α, subunit of the RNA polymerase III, in a tumor model. Human health and pathology. Université de Bordeaux, 2017. English. NNT : 2017BORD0822 . tel-02418602

HAL Id: tel-02418602
https://theses.hal.science/tel-02418602

Submitted on 19 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

Ecole doctorale : Science de la vie et de la santé

Spécialité : Génétique

Par Wiebke BRETTING

Study of RPC32α, subunit of the RNA polymerase III, in a tumor model

Sous la direction de : Dr. Hélène DUMAY-ODELOT

Soutenue le 11 décembre 2017

Membres du jury :

Prof. MERLIO, Jean-Philippe	*Professeur, Université de Bordeaux*	Président
Dr. MEGGETTO-PRADELLE, Fabienne	*Directrice de recherche, CRCT*	Rapporteur
Dr. ACKER, Joël	*Chercheur, CEA Saclay*	Rapporteur
Dr. DUMAY-ODELOT, Hélène	*Maître de conférence, Université de Bordeaux*	Examinateur
Prof. TEICHMANN, Martin	*Professeur, Université de Bordeaux*	Invité

Abstract

The RNA polymerases are key players of transcription. Eukaryotes have three RNA polymerases (I, II and III). The RNA polymerase III (Pol III) has 17 subunits, one of which exists in two alternative forms: RPC32α and RPC32β. Only one of the two forms can be integrated into the enzymes, thus generating either Pol IIIα or Pol IIIβ. While RPC32β is found in all somatic cells, RPC32α is expressed in stem cells and tumor cells. To date nothing is known of their respective roles.

Breast cancer is one of the major public health problems, as it is the most common cancer in women. Several types of breast cancers are distinguished, according to the presence or absence of hormonal receptors. Cancers that test negative for estrogen receptors, progesterone receptors and that do not overexpress the human epidermal growth factor receptor 2, are called triple-negative breast cancers. They tend to have a poor prognosis, due to the aggressive nature of the cancer and the lack of targeted therapies.

To study the role of RPC32α, a tumor model needed to be identified. In collaboration with Jean-Paul Feugeas (INSERM UMR 1098) a transcriptomic study was performed on 2627 clinical breast tissue samples. The study showed that RPC32α was overexpressed in triple-negative breast cancer, whereas RPC32β was overexpressed in normal tissue. A study on six breast cancer cell lines and one non-tumorigenic line confirmed the results of the transcriptomic study. The breast cancer model was thus validated.

A characterization of different breast cancer cell lines showed that other Pol III subunits were not overexpressed in triple-negative breast cancer. The overexpression of RPC32α was therefore not a mere consequence of a Pol III hyperactivity. An analysis of the transcripts synthesized by Pol III showed that overall the Pol III transcript levels were elevated in triple-negative breast cancer compared to other breast cancer subtypes.

In order to study the role of RPC32α in tumorigenesis, several RPC32α knock-out cell lines were created using CRISPR-Cas9. The loss of RPC32α did not induce an increase in transcription of the RNAs of RPC32α or RPC32β. This shows that no feed-back loop exists for RPC32α and that the two homologues are not co-regulated. Various Pol III transcripts showed decreased expression levels in the knock-out cell lines. Yet not all transcripts were

reduced in the absence of RPC32α. This indicates that some sort of transcription specificity must exist for Pol IIIα and Pol IIIβ.

The knock-out cell lines did not show any alterations in their phenotype or growth rates. However, in soft agar assays the knock-out cell lines produced 85% less colonies than the mother cell line. This proves that RPC32α is necessary for tumorigenic growth *in vitro*.

To find out if RPC32α was also necessary for tumorigenic growth *in vivo*, knock-out and wild type cells were injected into mice. The mice grafted with knock-out cells showed a slowed onset of tumor growth. After six weeks, the mice injected with knock-out cells had tumors half the size of the mice injected with wild type cells. The primary tumor was ablated and mice were tracked for metastasis. Four weeks later, mice injected with RPC32α knock-out cells had 100 times less metastasis than the control group. These results show that RPC32α is necessary for tumorigenic growth *in vitro* and *in vivo.* The protein seems also to be implicated in the formation of metastasis, which are one of the greatest problems in cancer treatment today.

Résumé

Les ARN polymérases sont des acteurs indispensables de la transcription. Chez les eucaryotes il existe trois ARN polymérases (I, II et III). La ARN polymérase III (Pol III) possède 17 sous-unités, dont une qui existe sous deux formes: RPC32α et RPC32β. Seulement une des deux formes peut être intégrée dans la Pol III, créant ainsi deux polymérases différentes Pol IIIα et Pol IIIβ. Alors que RPC32β est présent dans les cellules somatiques, RPC32α est exprimé surtout dans des cellules souches et des cellules tumorales. Aujourd'hui rien n'est connu sur leurs rôles respectifs.

Le cancer du sein est un problème majeur de santé publique car c'est le cancer féminin le plus fréquent. Plusieurs types de cancer du sein sont identifiés selon la présence ou absence de certains récepteurs hormonaux. Des cancers qui testent négative pour le récepteur d'œstrogène et de progestérone et qui ne surexpriment pas le récepteur pour les facteurs de croissance épidermiques humains 2 (HER2) sont appelés triple-négative. Ils ont un pronostique peu favorable, due à l'agressivité de ce type de cancer et un manque de thérapie cibles.

Pour étudier le rôle de RPC32α il fallait identifier un model tumorale. En collaboration avec Jean-Paul Feugeas (INSERM UMR 1098) une étude transcriptomique a été fait sur 2627 échantillons cliniques de tissus de sein. L'étude montre que RPC32α est surexprimé dans les cancers triple-négative, alors que son homologue RPC32β est surexprimé dans les tissues normaux. Une analyse sur six lignées de cancer du sein et une ligné non-tumorale ont pu confirmer les résultats de l'analyse transcriptomique. Le modèle de cancer du sein a donc été validé.

Une caractérisation des différentes lignées de cancer du sein a démontré que d'autres sous-unités de la Pol III n'étaient pas surexprimées dans les cancers triple-négative. La surexpression de RPC32α n'était donc pas une conséquence d'une hyperactivité de la Pol III. Une analyse des transcrits synthétisé par la Pol III a montré que en générale les transcrits de la Pol III étaient plus fortement exprimé dans les cancers triple-négative que dans d'autres cancers.

Afin d'étudier l'implication de RPC32α dans les phénomènes de tumorisation, plusieurs lignées cellulaires dépourvues de RPC32alpha ont été créé utilisant la technique CRISPR-CAS9. L'absence de RPC32α n'a pas induit une augmentation de transcription ni de l'ARN de

RPC32α, ni de celle de RPC32β. Il n'existe donc pas de boucle de rétroaction pour RPC32α et les deux homologues ne sont pas co-régulés. Plusieurs, mais pas tous les transcrits synthétisé par la Pol III ont une expression fortement baissé dans les lignées mutants. Le fait que pas tous les transcrits ne soit affectés par la perte de RPC32α, indique qu'il existe une spécificité de transcription pour Pol IIIα et Pol IIIβ.

Les cellules des linges mutants ne présentaient pas de phénotype différent des cellules mères et la croissance était la même dans toutes les lignées. Par contre les tests de croissance en agar-mou ont révélé que les lignées mutants formaient 85% de moins de colonies, indiquant que RPC32α est nécessaire pour la croissance tumorigénique *in vitro*.

Pour tester l'effet de la perte de RPC32α sur la croissance tumorigénique *in vivo*, des cellules mutants et des cellules mères ont été injecté dans des souris. Les souris greffées avec des cellules mutantes montrent un départ de tumorisation retardé. Au bout de six semaines elles avaient de tumeurs deux fois plus petit que les souris avec des cellules mères. Après ablation de la tumeur primaire, les souris ont été surveillées pour l'apparition de métastases. Quatre semaines plus tard les souris greffées avec des cellules mutantes avaient 100 fois moins de métastases que les souris contrôles. Ces résultats montrent que RPC32α est nécessaire pour la tumorisation *in vitro* et *in vivo*. La protéine semble surtout jouer un rôle dans la formation des métastases, qui sont un des problèmes majeurs dans le traitement des cancers.

Rapport du travail de thèse

Les ARN polymérases sont des acteurs indispensables de la transcription qui permettent de l'expression des gènes. Chez les eucaryotes il existe trois ARN polymérases nucléaires (I, II et III). L'ARN polymérase III (Pol III) est chargée de la transcription de petits ARN non-codants. Les ARN transcrits par la Pol III sont impliqués dans de nombreux processus cellulaires, tels que la transcription (ARN 7SK, éléments Alu), la transformation des ARN (ARNsn U6, RNase P, ARN MRP), la traduction (ARNr 5S, ARNts) et la maturation des protéines (7SL)... Une activité régulière de l'ARN polymérase III est donc essentielle pour la croissance et la survie de la cellule.

Depuis un certain temps, des études montrent que la dérégulation de la Pol III peut favoriser le cancer. Cependant aucun mécanisme exact n'a été identifié. Avec 17 sous-unités la Pol III est la plus complexe des trois polymérases nucléaires. Parmi ces 17 sous-unités il en existe une, RPC32, qui est présente sous deux formes indépendantes : RPC32α et RPCβ. Une seule de ces deux formes peut être intégrée dans la Pol III, créant ainsi deux polymérases différentes Pol IIIα et Pol IIIβ. Alors que RPC32β est présente dans les cellules somatiques, RPC32α est exprimée surtout dans les cellules souches (Wong *et al.*, 2011) et dans des cellules tumorales (Haurie *et al.*, 2010). Aujourd'hui rien n'est connu sur leurs rôles respectifs. Cette étude cherche à identifier le rôle de RPC32α dans un modèle tumoral et à élucider son potentiel comme cible pour des traitements anti-cancéreux.

Le cancer du sein est un problème de santé majeur. Selon l'organisation mondiale de la santé, il est le deuxième cancer le plus fréquent au monde. C'est le cancer du poumon qui détient la première place. Plusieurs sous-types de cancer du sein ont été décrits. Trois grandes catégories moléculaires sont identifiées : luminal, HER2 positive ($HER2^+$) et triple-négative. La classification se fait selon la présence ou l'absence de récepteurs hormonaux, notamment des récepteurs d'œstrogène, de progestérone ainsi que le récepteur du facteur de croissance humain (HER). Le cancer de sein dit triple-négative est testé négatif pour ces trois récepteurs.

Les différents types de cancer de sein présentent des pronostiques très variés. Les cancers triple-négative ont le pronostique le moins favorable. D'un coté, ce sont des cancers extrêmement agressifs, de l'autre, aucun traitement ciblé n'existe pour ces cancers. Il est

donc de la plus haute importance d'identifier une cible thérapeutique, pour mieux combattre les cancers triple-négatives.

Identification d'un modèle tumoral

Jusqu'ici le fonctionnement de la protéine RPC32α a été analysé dans des fibroblastes transformés (Haurie *et al.*, 2010) et des cellules souches (Wong *et al.*, 2011). Aucune étude n'avait caractérisé le rôle de RPC32α dans des cellules tumorales. Dans un premier temps, il fallait donc identifier un modèle tumoral approprié. Une recherche *in silico* a fait ressortir le cancer du sein comme modèle potentiel. En collaboration avec Jean-Paul Feugeas (INSERM UMR 1098), 2627 échantillons cliniques de tissus du sein ont été analysés par génie bioinformatique. L'expression du gène POLR3G, codant pour RPC32α, est fortement corrélée à des cancers du type triple-négatifs. Alors que l'expression du gène POLR3GL, codant pour RPC32β, est corrélée avec le tissu sain.

Au laboratoire, sept lignées cellulaires ont été mise en culture : trois lignées triple-négatives (MDA-MB231, BT-549 et MDA-MB468), deux lignées luminales (BT-474 et MCF7), une ligné $HER2^+$ ainsi qu'une lignée non-tumorale immortalisée (MCF-10A) comme contrôle. L'étude de ces lignées a montré que RPC32α est surexprimée dans les trois lignées triple-négatives, mais pas dans les autres lignées de cancer, aussi bien au niveau d'ARN que de la protéine. Par contre pour RPC32β le niveau d'ARN est le plus élevé dans la lignée non-tumorale. Au niveau protéique l'expression était variable, avec les niveaux les plus élevés dans les lignées MCF-10A et MDA-MB231.

Ces résultats sont en accord avec les données cliniques observées lors de l'étude bioinformatique. Le modèle de cancer de sein a donc été validé pour l'étude de RPC32α.

Caractérisation de RPC32α dans des lignées de cancer du sein

Comme une forte activité de l'ARN polymérase III a souvent été décrite dans des cellules cancéreuses, il était possible que la surexpression de RPC32α ne soit qu'un écho de cette surexpression. Une analyse d'autres sous-unités de la Pol III a montré qu'une légère surexpression de la Pol III a été observée dans toutes les lignées de cancer du sein. Mais seul RPC32α est surexprimée spécifiquement dans les cellules du type triple-négatif. Ces

résultats montrent que la surexpression de RPC32α n'est pas une suite de l'hyperactivité de la Pol III.

Pour obtenir de premières indications sur l'activité de RPC32α, différents transcrit de la Pol III ont été analysés dans chaque lignée cellulaire. De manière générale tous les transcrits analysés sont plus fortement exprimés dans les lignées triple-négatives comparé au contrôle non-tumorigénique. Les autres lignées montrent une expression plus faible que la lignée contrôle. Cependant la surexpression dans les lignées triple-négatives ne suit pas exactement le même profil que la surexpression de POLR3G. Il n'est alors pas possible d'en tirer des conclusions sur un possible rôle de RPC32α.

Création d'une lignée RPC32α knock-out *via* CRISPR-Cas9

Pour mieux pouvoir étudier RPC32α, il a été décidé de construire une lignée dépourvue de la protéine RPC32α (*knock-out* (KO)). Ainsi en comparant la lignée mère avec la lignée KO, il serait possible d'identifier des processus cellulaires dans lesquels RPC32α est impliquée. La lignée MDA-MB231 a été choisie comme lignée mère, car dans tous les tests précédents, elle s'est avérée comme bon représentant des lignées triple-négatives.

Pour créer la lignée *knock-out*, la méthode CRISPR-Cas9 a été choisie. L'ARN guide a été sélectionné de manière à ce que la coupure se fasse juste après le codon d'initiation de la transcription. La Cas9 a ensuite induit une coupure double brin de l'ADN. La coupure été réparée par la cellule *via* le mécanisme de réparation par jonction d'extrémités non-homologues (NHEJ). Ce mécanisme de réparation d'ADN est propice à l'erreur et des délétions/insertions sont souvent introduites. Ces mutations vont provoquer un déplacement du cadre de lecture dans le gène ciblé et vont ainsi abolir l'expression de la protéine.

Les cellules MDA-MB231 ont été transfectées avec un plasmide contenant les séquences de l'ARN de la Cas9. Après une sélection de trois jours avec la puromycine, les cellules ont été diluées à différentes concentrations dans des plaques 96 puits. La dernière concentration qui a donné des cellules a été utilisée pour la suite des analyses. Les cellules ont été expansées pour pouvoir extraire l'ADN. La zone de la coupure a été amplifiée par réaction en chaîne de la polymérase. Les amplicons ont été analysés sur gel acrylamide. Des candidats prometteurs

ont été envoyés à séquencer. Seul les candidats qui avaient un profil de séquençage avec des pics distincts ont été choisis pour la suite. Par immunodétection sur membrane, il a aussi été vérifié que la Cas9 ne s'était pas intégrée dans le génome de la cellule. Finalement les cellules ont été testées pour l'expression de RPC32α. Trois clones ont pu être identifiés qui n'expriment plus la protéine et qui sont donc RPC32α *knock-out*.

Caractérisation des lignées RPC32α *knock-out*

Les lignées RPC32α KO ont un niveau d'expression du gène POLR3G qui est similaire à celui de la cellule mère. Ce qui montre que la cellule ne détecte pas la perte de l'expression protéique de RPC32α, ou au moins, n'essaie pas de contrebalancer la perte de la protéine par une augmentation de la transcription. Il n'existe donc pas de boucle de rétroaction positive pour RPC32α. Une analyse de RPC32β au niveau de l'ARN et de la protéine a montré que l'homologue de RPC32α n'est pas surexprimée suite à la perte de RPC32α. Ceci indique que les deux homologues ne sont pas co-régulés. Ces résultats sont particulièrement intéressants, car cette étude est la première étude sur RPC32α, dans laquelle des cellules KO sont utilisées. Toutes les études précédentes se sont servies de modèles où l'expression de RPC32α était diminuée avec des ARN interférents. Jusqu'ici, la possibilité existait qu'une faible quantité de RPC32α suffisait pour empêcher une augmentation de RPC32β. Avec les cellules KO, il a pu être établi que même en absence totale de RPC32α, RPC32β n'est pas surexprimée. De la même manière la perte de RPC32α n'augmente pas l'expression d'autres sous-unités de la Pol III, ni au niveau de l'ARN, ni de la protéine.

Une analyse de plusieurs transcrits de la Pol III a montré que certains transcrits ont une forte baisse d'expression dans les cellules KO. C'est surtout le cas des deux ARNt$^{Met\ i}$ et ARNt$^{Met\ e}$, ainsi que l'ARN 7SL, ont des expressions plus faibles que la ligne mère. D'autres ARN ou l'expression avait diminué étaient l'ARNr 5S et l'ARN MRP. L'ARN BC200, par contre, est exprimé à des niveaux comparables à celui de la ligne mère. Ces résultats montrent qu'il existe un groupe de transcrits de la Pol III qui est plus affecté par la perte de RPC32α que d'autres. Cela pourrait indiquer que la Pol IIIα transcrit un autre ensemble d'ARN que la Pol IIIβ. Le rôle de RPC32α pourrait donc inclure de pousser la Pol IIIα vers la transcription de certains ARN distincts, ces ARN pourraient ensuite être impliqués dans la tumorigénèse.

Malgré une baisse de certains ARN structuraux, tels que l'ARNr 5S out les $ARNt^{Met\ i}$ et $ARNt^{Met\ e}$, les cellules KO ont conservé le même phénotype que leur cellule mère. Aussi, aucun changement au niveau de la prolifération n'a été noté. Pour tester l'effet du KO sur la tumorigénicité des cellules, un test en agar mou a été fait. Dans ce test, les cellules sont suspendue dans un agar qui les empêche d'adhérer à une surface et de communiquer entre elles. Les cellules non-tumorales ne peuvent pas croître sous de telles conditions. Par contre, les cellules tumorales vont se diviser et former des colonies dans l'agar mou. Il a été observé que les cellules 32α KO forment en moyenne 85% moins de colonies que les cellules mères. Ce résultat montre que RPC32α est important pour la tumorigénèse. Ainsi la surexpression de RPC32α dans des cellules tumorales n'est pas une conséquence, mais une cause de la tumorisation.

Experience *in vivo*

Afin de vérifier si les observations faites *in vitro* étaient reproductibles *in vivo*, la lignée mère et un clone des cellules RPC32α KO ont été transduits avec un vecteur portant le gène de la luciférase. En collaboration avec Elodie Richard (INSERM U1218), dix souris par groupe ont été greffées avec soit des cellules mères, soit des cellules RPC32α KO. Les cellules ont été injectées dans le canal galactifère, afin de reproduire le plus possible le microenvironnement d'une tumeur de sein. Par luminescence l'évolution de la tumeur a pu être observée durant l'expérience.

Alors que les tumeurs des cellules mère croissent de semaine en semaine, il a été observé que les tumeurs des cellules RPC32α KO diminuent, avant de reprendre leur croissance. Ainsi, les tumeurs des cellules KO montrent un retard de croissance tumorigénique. Au bout de six semaines, les glandes mammaires contenant les tumeurs ont été enlevées. Les tumeurs de la lignée KO ont seulement la moitié de la taille que celles de la ligne mère. Ceci montre que RPC32α est important pour la croissance tumorale *in vivo*.

Après ablation de la tumeur primaire, les souris ont été observées pendant quatre semaines pour l'apparition de métastases. A la fin des quatre semaines, les souris avec des tumeurs de la lignée mère ont un niveau de métastases 100 fois plus élevé que les souris KO. RPC32α semble jouer un rôle important dans la formation des métastases. Il a été montré par cette étude que RPC32α est important pour la tumorigènese et d'autres études ont montré que

RPC32α est important pour le maintien de la pluripotence dans les cellules souches (Wong *et al.*, 2011), ainsi un rôle de RPC32α pour le maintien des cellules souches cancéreuses (CSC) pourrait être envisagé.

Pour tester cette hypothèse, les tumeurs primaires ont été analysées par immunohistochimie pour des marqueurs de prolifération, de CSC et de la transition épithéliale mésenchymale. Aucun des marqueurs analysés (Ki67, CD44, E-cadherin, vimentine) ne montre une différence entre cellules mères et cellules KO. Alors que ces résultats ne confirment pas un rôle de RPC32α dans le maintien des CSC, ils ne l'excluent pas non plus. Les CSC ne sont pas définies uniquement par la présence de CD44, et une transition épithéliale mésenchymale peut se faire sans changement d'expression de E-cadherin et vimentine. De plus, il est possible que les tumeurs se soient formées à partir de cellules qui avaient des mutations supplémentaires, ce qui les a sauvées de la perte de RPC32α. Ceci pourrait expliquer pourquoi l'évolution de la tumeur connaît d'abord une baisse (mort des cellules RPC32α KO) avant de reprendre la croissance (amplification des cellules autrement mutées). Dans ce cas, l'analyse des tumeurs ne serait pas représentative de la situation dans des cellules RPC32α KO.

Des travaux supplémentaires vont être nécessaires pour comprendre la fonction de RPC32α dans des cellules tumorales. Une étape importante sera de répéter l'expérience *in vivo* avec un autre clone KO de la ligné MDA-MB231, mais aussi avec d'autres cellules RPC32α KO créées à partir d'autres lignées triple-négatives. En préparation de cette étape, les deux lignées BT-549 et MDA-MB468 ont également été mutées par CRISPR-Cas9. La séquence de l'ARN guide a été changée, dirigeant la Cas9 à couper plus en aval du codon d'initiation. Ainsi tout risque d'un effet de lignée cellulaire ou d'effet secondaire de la Cas9 peut être éliminé.

Pour la lignée BT-549 un clone a été identifié qui n'exprime plus RPC32α et un deuxième qui montre une très faible expression. Ceci peut être dû, soit à la présence d'une cellule non mutée, soit à la présence d'un allèle sauvage, dans le cas d'une mutation hétérogène. Dans la lignée MDA-MB468 plusieurs clones ont été identifiés ou l'expression de RPC32α a été réduite, mais pas éliminée. Pour tous les clones ayant une expression faible, un nouveau traitement par CRISPR-Cas9 peut éliminer toute trace d'expression de RPC32α.

Régulation de RPC32α

Comme l'expression de RPC32α est importante pour la tumorigénèse, il est essentiel que son expression soit strictement régulée. Une analyse *in silico* a révélé qu'une séquence riche en guanines était présente dans l'intron 1 de POLR3G, qui avait le potentiel de former un G-quadruplex. Ces structures secondaires se créent quand des guanines se lient grâce à des appariements de base de type Hoogsten. Les guanines forment ainsi un plateau appelé quartet. L'empilement de plusieurs quartets constitue un G-quadruplex (G4).

Des analyses *in vitro* ont été effectuées, pour confirmer expérimentalement que la séquence identifiée par analyse *in silico,* était capable de former un G4. Les résultats montrent que la séquence est capable de se lier en G-quadruplex, qui est stable à des températures supérieures à 40° C. Il est donc possible que un G4 se forme dans des cellules au niveau de l'intron 1 de POLR3G. Le rôle d'une telle structure pourrait être de bloquer la transcription du gène. Si cette hypothèse est correcte, les cellules devraient montrer une surexpression de l'exon 1 de POLR3G, qui correspond à des transcrits tronqués, issus d'une transcription abortive. Une telle surexpression a été observée pour des lignées du type triple-négative, les mêmes cellules qui montrent une surexpression de POLR3G.

Il est possible que le G-quadruplex soit un régulateur négatif de la transcription et que les cellules avec une forte expression de POLR3G essaient de diminuer cette expression en stabilisant le G4. Une autre possibilité est que le G4 soit un point d'ancrage pour des activateurs de transcription. Ainsi, même si certains transcrits sont abortifs, à cause de l'inhibition stérique, le résultat global est une augmentation de l'expression génique. D'autres études seront nécessaires, pour mieux comprendre le rôle de ce G4.

Conclusion

Le but de cette étude était de caractériser le rôle de RP32α dans un model tumoral. Dans un premier temps, le modèle de cancer du sein a été établi au laboratoire et validé pour l'étude de RPC32α. Sa pertinence a été montrée en analysant non seulement des lignées cellulaires, mais aussi 2627 échantillons cliniques de tissus de sein. Le modèle de cancer du sein, établi par cette étude, peut désormais être utilisé pour des recherches futures.

De plus cette étude a augmenté les connaissances sur le fonctionnement et le rôle de RPC32α. Il a été établi que RPC32α est important pour la tumorisation *in vitro* et *in vivo*. De plus RPC32α semble jouer un rôle clé dans la formation des métastases. Ces résultats mettent en évidence le potentiel de RPC32α comme cible thérapeutique. Surexprimé dans les cancers du sein triple-négatifs, RPC32α pourrait servir à combattre cette forme agressive du cancer, qui manque pour l'heure d'un traitement spécifique.

Bien que le travail fourni par cette étude ait donné des résultats prometteurs, d'autres recherches seront nécessaires pour élucider le rôle de RPC32α. Ces futurs travaux pourront se baser sur des modèles et des outils créés par cette étude. Notamment les lignées RPC32α KO crées dans le cadre de cette thèse, seront d'une grande utilité. Ainsi, ce travail a, non seulement fait progresser les connaissances sur RPC32α, mais a aussi posé des fondements essentiels pour des recherches futures.

Acknowledgements

First of all I would like to thank the members of the jury, Prof. Jean-Philippe Merlio, Dr. Fabienne Meggetto-Pradelle et Dr. Joël Acker, for having taken the time to evaluate this work.

I would like to thank Hélène Dumay-Odelot for having entrusted me with this project. It has been great to be working with and learning from you. You have been much more for me that just my PhD supervisor and I will never forget that you let us store all our furniture in your living room. Furthermore, I would like to thank Martin Teichmann for accepting me into his team and the many scientific, politic and worldly discussions I enjoyed. To Stéphanie Durrieu-Gaillard I owe a great debt of gratitude for teaching me the numerous tricks needed to make an experiment work and to survive in the lab. One day we will find that special idea that will make us rich. I would also like to thank my fellow PhD students Camila Parrot and Emilie Rousseau for shared wisdom, fun and conversations.

I also would like to thank the members of the Unit 1212 for moral support and for making my time in the lab much more memorable. I would like to thank the Innis Team for shared coffee breaks and chocolats. Especially I would like to thank Britta Seip for endless conversations and for letting me know what it is like to live with a cat, Caroline Seefeldt for always looking on the bright side and for teaching me how to color proteins, Alba Herrero for her smile and kind words when needed and Natacha Perebaskine for always having an open ear.
All the unit members working in Carreire I would like to thank for welcoming us to the new site and answering my multiple “where is” and “where can I find” questions. I would like to thank Jaques Puyol for showing me where everything was and for being super patient. I would like to thank all those people who made lunch break more fun. And last but not least I would like to thank Kati Ba-Pierozzi for being super reactive and staying cool in all situations (even when confronted with a panicked PhD student).

But it was not just the people at work who have helped me to finish this work. I also want to give my deepest gratitude to my parents for having encouraged me to take up my studies again and to support me all along. I also want to thank my brother who was always there when I needed him most. And finally I would like to thank Jérôme, who lived through it all right next to me and who believed in me more than I did myself. Thank you!

Table of contents

List of figures 18

List of tables 19

Abbreviations 20

Chapter I – Introduction 23

1. Transcription by the RNA polymerase III 25
1.1. Transcripts of the human RNA polymerase III 26
1.2. 5S ribosomal RNA (5s rRNA) 26
1.3. Transfer RNAs (tRNAs) 27
1.4. U6 small nuclear RNA (U6 snRNA) 28
1.5. H1 RNA or RNase P RNA 28
1.6. MRP RNA 29
1.7. 7SL RNA 29
1.8. Vault RNAs 29
1.9. Y RNAs 30
1.10 7SK RNA 30
1.11. BC1 and BC200 RNA 31
1.12. Virus encoded RNAs 31
1.13. SINE and ALU 32
1.14. Other transcripts 33

2. The Pol III transcription apparatus 33
2.1. Promoter type I 34
2.2. Promoter type II 35
2.3. Promoter type III 35
2.4. Mixed promoters 36
2.5. Transcription termination 36
2.6. Transcription factor III A (TFIIIA) 37
2.7. Transcription factor III C (TFIIIC) 37
2.8. Transcription factor III B (TFIIIB) 41
2.9. snRNA activated protein complex (SNAPc) 42
2.10. Epigenetic factors 43

3. The RNA polymerase III 44
3.1. The core 47
3.2. The stalk 48
3.3. The heterodimer 48
3.4. The heterotrimer 49
3.5. The subunits RPC32α and RPC32β 51
3.5.1. Evolution 51
3.5.2. Structure 52
3.5.3. Expression 53
3.5.4. Function 54
3.5.5. Regulation 55

4. Pol III and cancer 56

5. Breast cancer *59*
5.1. Risk factors 60
5.2. Diagnosis 61
5.3. Molecular subtypes of breast cancer 63
5.3.1. Luminal cancers 63
5.3.2. HER2 positive cancers 64
5.3.3. Triple-negative breast cancers 65
5.3.3.1. Basal cancers 66
5.3.3.2. Claudin-low and MBC cancers 66
5.3.3.3. Normal-like cancers 66
5.3.4. Molecular apocrine cancers 67
5.4. Breast cancer stem cells 68
5.4.1. Breast cancer stem cell markers 69
5.4.2. Stem-cellness and therapy resistance 69
5.4.3. Epithelial-to-mesenchymal transition 70

6. Purpose of this study *72*

Chapter II – Materials & Methods **73**

Materials **75**

1. Cell lines and culture conditions *76*

2. Bacterial strains *76*

3. Plasmids and expression vectors *76*

4. Primers and oligonucleotides *77*
4.1. Primers for RT-qPCR 77
4.2. Primers for PCR 77
4.3. Oligonucleotides 78

5. Antibodies *78*

Methods **78**

1. Creation of competent bacteria *78*

2. Bacterial transformation *79*

3. Plasmid extraction *79*
3.1. Miniprep 79
3.2. Midiprep 80

4. CRISPR *80*
4.1. Design 80
4.2. Annealing and phosphorylation 81
4.3. Digestion and ligation 82
4.4. Cellular transfection 82

4.5. Identification of potential knock-out cell lines 82

5. DNA extraction *83*
5.1. DNA extraction from cell culture dishes 83
5.2. DNA extraction from 96 well plates 83

6. RNA extraction *84*

7. RNA quality control *84*

8. Reverse transcription *85*

9. RT-qPCR *85*

10. Protein extraction *86*

11. Protein quantification *86*

12. Western blots *86*
12.1. Protein separation by gel electrophoresis 86
12.2. Protein transfer on to a nitrocellulose membrane 87
12.3. Blocking and antibody incubation 87

13. MTT assay *87*

14. Wound healing assay *88*

15. Soft agar assay *88*

16. Transduction of cells with a luciferase gene *88*

17. Test for luminescence *89*

18. Orthotopic mouse xenografts *89*

19. Histological analysis of tumor tissue *89*

20. Absorbance spectroscopy *90*
20.1. Isothermal Differential Absorbance Spectra (IDS) 90
20.2. UV melting analysis 90

21. Statistics *90*

Chapter III – Results **91**

1. Identification of an appropriate tumor model *93*

2. Characterization of RPC32α in different breast cancer cell lines *98*
2.1. RPC32α is overexpressed in triple-negative breast cancer on an RNA and protein level 98
2.2. No other POLIII subunits are overexpressed in triple-negative breast cancer 100

2.3. RPC32α and Pol III transcripts 102

3. Creation of an RPC32α knock-out cell line *103*
3.1. Cutting POLR3G with CRISPR-Cas9 104
3.2. Characterization of RPC32α knock-out cell lines 106
3.2.1. No feed-back loop exists in the regulation of RPC32α 106
3.2.2. RPC32α and RPC32β are not co-regulated 107
3.2.3. The knock-out of RPC32α does not lead to a change in overall Pol III levels 108
3.2.4. RPC32α knock-out cells show reduced transcription of some, but not all Pol III transcripts 109
3.2.5. RPC32α knock-out cell lines do not have an altered phenotype or rate of proliferation 110
3.2.6. Unchanged migration capacity in RPC32α knock-out cells 111
3.2.7. RPC32α is important for tumorigenic growth *in vitro* 111
3.3. Tumorigenicity of RPC32α knock-cell lines *in vivo* 113
3.3.1. Creation of luminescent cell lines to follow tumor development *in vivo* 113
3.3.2. Loss of RPC32α leads to reduced tumor growth *in vivo* 115
3.3.3. The RPC32α knock-out cells generate significant less metastasis 116
3.3.4. Histological analysis of the primary tumor 117
3.3.5. Histological analysis of the metastases 121

4. Creation of new RPC32α knock-out cell lines *122*

5. Regulation of RPC32α *124*

Chapter IV – Discussion **131**

1. What role for RPC32α and RPC32β? *133*

2. How could Pol III transcripts promote tumorigenic growth? *136*
2.1. tRNAs 136
2.2. tRNA derived fragments (tRF) 139
2.3. Other Pol III transcripts 140
2.4. Roles for RPC32α beyond transcription 142

3. Regulation of RPC32α *145*
3.1. G-quadruplexes 145
3.2. miRNAs 148

4. Is RPC32α related to cancer stem cells? *150*

5. Could RPC32α become a new drug target? *153*

6. Outlook *155*

Chapter V – Bibliography **157**

List of figures

Figure 1: A few of the cellular roles of Pol III transcripts. 26
Figure 2: Promoter types of the RNA polymerase III. 34
Figure 3: Promoters and transcription factors of the RNA polymerase III (Pol III). 38
Figure 4: Structure of the TBP:DNA interaction and the Bdp SANT domain. 42
Figure 5: Histone modifications that influence transcription by Pol II and Pol III. 44
Figure 6: Cryo-EM structure of the eukaryotic RNA polymerase I, II and III. 46
Figure 7: Transcription model for the yeast RNA polymerase III. 47
Figure 8: Comparison of the closed clamp formation in Pol III versus Pol II and Pol III. 48
Figure 9: Architecture of the Pol III specific heterotrimer. 50
Figure 10: Number of the different POLR3G-like genes in different organisms. 52
Figure 11: Structure of the RPC32β- RPC62 complex. 53
Figure 12: The regulatory network of Pol III. 57
Figure 13: 5-year survival rate of breast cancer patients by country. 59
Figure 14: Anatomy of the female breast. 61
Figure 15: Evolution from a normal milk duct to an invasive ductal carcinoma. 62
Figure 16: Hazard rate of reoccurrence. 65
Figure 17: Two models of heterogeneity in tumors. 68
Figure 18: The Epithelial-to-mesenchymal-transistion (EMT). 71
Figure 19: The role of EMT and MET in the formation of metastasis. 72
Figure 20: Expression levels of Pol III subunits and transcription factors in different molecular breast cancer subtypes. 94
Figure 21: Correlation of Pol III subunits with different breast cancer subtypes. 95
Figure 22: Correlation of Pol III subunits that have homologs in Pol II with different breast cancer subtypes. 96
Figure 23: Correlation of Pol III transcription factors with different breast cancer subtypes. 97
Figure 24: Expression levels of POLR3G and POLR3GL. 99
Figure 25: Expression levels of RPC32α and RPC32β. 100
Figure 26: Expression levels of other Pol III subunits. 101
Figure 27: Expression levels of different Pol III transcript. 102
Figure 28: Mutating POLR3G using the CRISPR-Cas9 system. 104
Figure 29: Creation of an RPC32α knock-out cell line. 105
Figure 30: Expression levels of RPC32β and POLR3GL in the RPC32α knock-out cell lines. 107
Figure 31: Expression levels of other Pol III subunits in RPC32α knock-out cells. 108
Figure 32: Expression levels of different Pol III transcripts in RPC32α knock-out cells. 109
Figure 33: Phenotype and proliferation rate of RPC32α knock-out cells. 110
Figure 34: Wound-healing assay to assess migration capacity of RPC32α knock-out cells. 111
Figure 35: Soft agar assay to assess anchorage-independent growth in RPC32α knock-out cells. 112
Figure 36: Transduction of cell lines with luciferase containing vector. 114
Figure 37: Tumor development *in vivo*. 115
Figure 38: Metastasis in the wild type and knock out group. 117
Figure 39: Histological analysis of the primary tumors. 118

Figure 40: Histological analysis of the primary tumors. 119
Figure 41: Immunohistochemical staining of the primary tumor. 120
Figure 42: Hematoxylin and eosin staining of affected organs. 121
Figure 43: Creation of new RPC32α knock-out cell lines. 123
Figure 44: Potential G-quadruplex structure in POLR3G. 125
Figure 45: G-quadruplex sequences tested. 126
Figure 46: *In vitro* analysis of different G4 sequences. 127
Figure 47: *In vitro* analysis of mutated G4 sequences. 128
Figure 48: Expression analysis of the first exon of POLR3G. 129
Figure 49: RPC32α and RPC32β occupy largely identical sites in IMR90 fibroblasts. 134
Figure 50: TERT increases RPC32α binding to the target genomic regions. 137
Figure 51: Different types of tRNA derived fragments (tRF) and their implications in cancer. 139
Figure 52: RNAs altered after the knock-down of RPC32α in human embryonic stem cells. 141
Figure 53: Chromatin opening and the relationship between Pol II and Pol III. 143
Figure 54: Potential effects of a G-quadruplex on transcription. 147
Figure 55: MiRNA regulating the expression of POLR3G. 149
Figure 56: Impact of traditional versus cancer stem-cell directed therapy. 154

List of tables

Table 1: Overview of promoters found in different genes transcribed by the human Pol III. 33
Table 2: Homologies between yeast and human TFIIIC subunits and their respective function. 39
Table 3: The RNA polymerases and their subunits across species. 45
Table 4: Molecular subtypes of breast cancer. 63
Table 5: Cell lines used in this work and the breast cancer subtypes they represent. 98

Abbreviations

ALDH1:	Aldehyde dehydrogenase 1
BDP1:	B double prime 1
BER:	Base excision repair
BRCA 1/2:	Breast cancer gene ½≤
BRF1/2:	TFIIIB related factor 1/2
CdK9:	Cyclin dependent kinase
cDNA:	Complementary DNA
ChIP-seq:	Chromatin Immunoprecipitation sequencing
CRISPR:	Clustered Regularly Interspaced Short Palindromic Repeats
DNA:	Deoxyribonucleic acid
DSE:	Distal sequence element
EMT:	Epithelial-mesenchymal transition
ER	Estrogen receptor
ETC:	Extra TFIIIC loci
G4:	G-quadruplex
GTF:	General transcription factor
HER2:	Human epidermal growth factor receptor 2
HEXIM:	Hexamethylene bisacetamide inducible protein
hnRNP:	Heterogeneous nuclear ribonucleoproteins
HR:	Homologous recombination
ICR:	Internal control region
IE:	Intermediate element
IP:	Immunoprecipitation
KO:	Knock-out
KRAS:	Kirsten rat sarcoma
LUCA:	Last universal common ancestor
mRNA:	Messenger RNA
MET:	Mesenchymal-epithelial transition
MTT:	3-(4.5-dimethylthiazol-2-yl)-2.5-diphenyltetrazolium bromide
MVP:	Major vault protein
NDR:	Nucleosome depleted region
NEP:	Nuclear encoded polymerase
NF1:	Nuclear factor 1
NHEJ:	Non-homologous end joining
OD:	Optical density
PARP:	Poly-ADP-ribose-polymerases
PBP:	PSE-binding protein
PCR:	Polymerase chain reaction
PEP:	Plastid encoded polymerase
PKR:	Protein Kinase R
Pol I-IV:	RNA polymerase I-IV
PR:	Progesterone receptor
PSE:	proximal sequence element
p-TEFb:	Positive transcription elongation factor
PTF:	PSE transcription factor
qPCR:	Quantitative polymerase chain reaction
rDNA:	Ribosomal DNA
Rb:	Retinoplastoma susceptibility gene
RdRP:	RNA dependent RNA polymerase

RMRP:	RNase mitochondrial RNA processing
RNA:	Ribonucleic acid
RNAP:	RNA polymerases
RoRNP:	Ro ribonucleoprotein
rRNA:	Ribosomal RNA
RT:	Reverse transcription
SINE:	Short interspersed elements
SNAPc:	snRNA activating protein complex
snRNA:	Small nuclear RNA
snRNP:	Small nuclear ribonucleoproteins
SRP:	Signal recognition particle
SSB:	Single stranded break repair
TBP:	TATA-binding protein
TE:	Transposable element
TEP1:	Telomerase-associated protein
TF:	Transcription factor
TFIIIA:	Transcription factor A of the RNA polymerase III
TFIIIB:	Transcription factor B of the RNA polymerase III
TFIIIC:	Transcription factor C of the RNA polymerase III
TNBC:	Triple negative breast cancer
TRF1:	TBP-related factor 1
tRF:	tRNA-derived fragment
tRNA:	Transfer RNA
TSS:	Transcription start site
VA RNA:	Virus associated RNA
VEGF:	Vascular endothelial growth factor
VEGFR-2*:*	Vascular endothelial growth factor receptor 2
vPARP:	Vault poly(AD-P) ribose polymerase
WT:	Wild type

Chapter I

Introduction

For 15 years now the laboratory of Gene Regulation and Tumor Research has helped to elucidate the role of the RNA polymerase III (Pol III) in the onset of cancer. Significant discoveries were made using transformed fibroblasts, notably the existence of Pol III subunit RPC32α (Haurie et al. 2010). One aim of the present study was to establish a tumor model based on cancer cell lines. This model is to serve not only for this work, but also for future studies in the laboratory, so that the activity of RPC32α and Pol III may be analyzed in a more realistic tumor environment.

1. Transcription by the RNA polymerase III

The genome of a cell contains all the information a cell needs to grow, divide and multiply. In order to access this information however, the DNA has to be transcribed into RNA. This is done by the RNA polymerase. Given the evolutionary pressure, it is not surprising that RNA polymerases show a highly conserved structural framework across species (reviewed in Werner, 2008). The simplest form can be found in bacteria. It is probably also the form that resembles the most the *last universal common ancestor* (LUCA) of bacteria, archaea and eukaryotes. The bacterial RNA polymerase possesses four subunits that form the core enzyme: β, β', α and ω (Burgess 1969). While this core enzyme can perform transcription, it needs a fifth factor to form the holoenzyme. This so called σ-factor recruits the RNA polymerase to defined promoters and thus ensures specific transcription (Burgess et al. 1969).

RNA polymerases have become more complex during evolution, but homologs to the bacterial core enzymes can be found across species. Eukaryotes typically have three nuclear RNA polymerases (Pol) -Pol I, Pol II and Pol III (Roeder & Rutter 1969). In 1986 another RNA polymerase was identified in the mitochondria of yeast (Greenleaf *et al.,*1986). Surprisingly the mitochondrial RNA polymerase did not resemble the bacterial RNA polymerase, but rather the RNA polymerases of the bacteriophage family T3/T7 (Masters *et al.*, 1987). The human mitochondrial RNA polymerase was identified ten years later and like its yeast counterpart it is formed of only one single subunit (Tiranti et al. 1997).

Plants possess four additional RNA polymerases. Two are found in the chloroplasts, one is bacterial-like and plastid encoded (PEP), while the other is phage-like and nuclear encoded (NEP) (Börner et al. 2015). Furthermore, plants have two non-essential polymerases – Pol IV and Pol V (Wierzbicki *et al.*, 2008; Ream *et al.*, 2009), which play a role in RNA mediated DNA methylation (Movahedi et al. 2015).

Each RNA polymerase transcribes a specific subset of RNAs. RNA polymerase I (Pol I) is unique among the nuclear RNA polymerases, as it synthesizes only one single transcript: the precursor rRNA. The RNA polymerase II (Pol II) transcribes all mRNAs and a small set of non-coding RNAs such as small nuclear RNAs, small nucleolar RNAs and micro-RNAs. Other non-coding RNAs that are essential for

cell survival are transcribed by the RNA polymerase III (Pol III) (reviewed in Schramm and Hernandez, 2002 and Wild and Cramer, 2012).

1.1. Transcripts of the human RNA polymerase III

The transcripts synthesized by the RNA polymerase III (Pol III) are small non-coding RNAs that are involved in many cellular processes, such as regulation of transcription, RNA processing and translation (figure 1). The deregulation of Pol III transcripts has been linked to severe diseases such as cancer and Alzheimer's disease (White, 2004; Mus *et al.*, 2007; Lee, 2015). Furthermore, it has been shown that enhanced Pol III transcription is necessary for tumorigenic growth (Johnson, *et al.,* 2008).

While the discovery of the first Pol III transcripts dates back several decades, new potential Pol III loci are still being discovered and their function remains to be discovered (Leśniewska & Boguta 2017; Lee 2015; Dieci et al. 2007) .

1.2. 5S ribosomal RNA (5s rRNA)

The eukaryotic ribosome is composed of two ribosomal RNA subcomplexes, the large or 60S subunit and the small or 40S subunit. The small subunit is formed by one single RNA (18S), whereas the larger

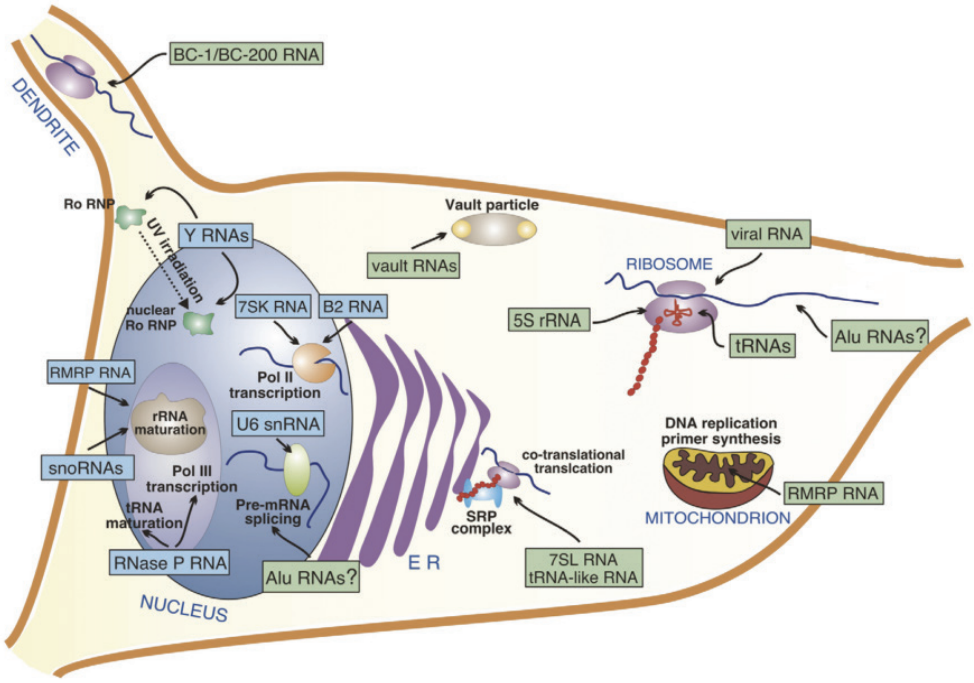

Figure 1: A few of the cellular roles of Pol III transcripts. Transcripts that are active in the nucleus are colored in blue, those that play a role in the cytoplasm are green. Details about the function of each transcript can be found in the main text.

Image adapted from Dieci *et al.*, 2007

subunit contains three RNA species (5S; 5.8S and 28S). The 18S, 5.8S and 28S RNAs are transcribed as one single pre-rRNA transcript unit by the RNA polymerase I. Only the 5S RNA is synthesized by the RNA polymerase III.

In most eukaryotes the genes coding for the 5S rRNA are located in one large cluster that is independent from the rDNA transcribed by Pol I (reviewed in Goodfellow and Zomerdijk, 2012); however, in some eukaryotes, such as *Saccharomyces cerevisiae*, the 5S rRNA are located directly adjacent to the unit transcribed by Pol I (Bell et al. 1977). In the human genome there exist up to 2000 copies of the 5S rDNA. The majority of them are located in a single tandem array on chromosome 1, but 200-300 copies of the 5S gene are dispersed throughout the genome (Steffensen *et al.*, 1974; Soerensen and Frederiksen, 1991; Matera and Ward, 1992; Cooper, 2000).

For the biogenesis of one ribosome exactly one copy of each of the four different RNA species (5S, 5.8S, 18S and 23S) is needed. Given that three of them are transcribed by Pol I and one by Pol III, it is necessary to coordinate transcription levels between the two polymerases. A number of proteins are an integral part of the ribosome and their mRNAs are transcribed by Pol II, therefore the transcription levels of all of these RNAs need to be concerted. Indeed it was shown that cells under stress have simultaneous decrease in mRNAs encoding r-proteins and 35S rRNA (Mizuta & Warner 1994; Powers & Walter 1999), r-protein mRNAs and 5S rRNA (Li et al. 2000) or 35S rRNA and 5S rRNA (Clarke et al. 1996; Zaragoza et al. 1998). Ultimately it was shown that a decrease in Pol I transcription, leads to a decrease in Pol II transcribed r-protein mRNA and in the Pol III transcribed 5S rRNA (Laferté et al. 2006). This proves that the three RNA polymerases act in a concerted fashion to regulate ribosome production.

Although the exact mechanisms behind this simultaneous deregulation have not been elucidated, there is some evidence that spatial proximity between rDNA is a key factor. For even if the different rDNAs are distant on a linear genomic map, they seem to be in close proximity in the nucleus. Due to the folding and coiling of the DNA in the nucleus the 5S gene can be found at or near the nucleolus, site of the Pol I rRNA transcription (Haeusler & Engelke 2006; Montijn et al. 1999). This spatial proximity could allow for a coordinated transcription of all rRNAs.

1.3. Transfer RNAs (tRNAs)

Transfer RNAs serve as an adapter from the RNA to the protein level. Typically, tRNAs are 76 to 90 nucleotides long and have a cloverleaf structure, characterized by three stem loop structures and one acceptor stem. The loop opposite of the acceptor stem contains the anticodon, a sequence of three nucleotides that will code for a specific amino acid. The genetic code is degenerate that is why there are a several isoacceptors for each of the 20 amino acids and the stop codon. The number of

tRNAs genes varies from species to species, *Saccharomyces cerevisiae* has 275 and humans possess a predicted 522 tRNA genes plus 109 tRNA derived pseudogenes (Goffeau *et al.*, 1996; Canella *et al.*, 2010).

Like ribosomes, tRNAs are essential for translation. Recent studies indicate that tRNAs are not merely the link between mRNA and proteins, but that changes in the tRNA pool can influence mRNA translation and thereby protein availability (Grewal 2014; Pavon-Eternod et al. 2013). Furthermore, it has been shown that both pre- and mature tRNAs can undergo endonucleatic cleavage, which will produce short RNA fragments. These tRNA-derived fragments (tRFs) are implicated in a number cellular processes, such as apoptosis, protein synthesis control and RNA interference (Soares & Santos 2017).

1.4. U6 small nuclear RNA (U6 snRNA)

In eukaryotic cells a complex of 5 small nuclear ribonucleoproteins (snRNPs), called the spliceosome, ensures accurate splicing of mRNAs. The 5 corresponding small nuclear RNAs (snRNAs) are U1 snRNA, U2 snRNA, U4 snRNA, U5 snRNA and U6 snRNA. While four of them (U1, U2, U4 and U5) are transcribed by the RNA polymerase II, the fifth snRNA (U6) is transcribed by Pol III. The human U6 snRNA was first discovered in 1980 (Daskal *et al.*, 1980) and unlike other splicosomal RNAs it shows remarkable similarity in size, sequence and structure to its yeast homologue (Brow & Guthrie 1988). Another particularity of U6 among the splicing snRNAs is that its entire maturation takes places in the nucleus (Kunkel et al. 1986). Humans possess at least nine copies of the U6 gene, which are dispersed thought the genome, but only 5 of them have to be proven functional (Domitrovich & Kunkel 2003).

1.5. H1 RNA or RNase P RNA

The ribonuclease P is an endoribonuclease that is involved in the maturation of tRNA. In eukaryotes it is composed of an RNA part (H1 RNA) and nine to ten associated proteins. Surprisingly, it is not the proteins that cut the tRNA, they merely have a catalytic function. Indeed, the RNA chain of the ribonuclease is capable of cleaving tRNA without the help of proteins. For this discovery Sidney Altman earned the 1989 Nobel Prize, as it proved that RNAs could function as enzymes, a major prerequisite in the model of the RNA world. Altman had performed his works using bacteria, but it took over two decades to prove that eukaryotic RNase P as well can cleave RNA in the absence of proteins (Kikovska *et al.*, 2007).

Besides its role in tRNA maturation, RNase P also functions as a transcription factor to RNA polymerase I and III (Reiner et al. 2006; Reiner et al. 2008). While it was shown that depletion of RNase P or the targeted cleavage of H1 RNA led to deficiencies in Pol I and Pol III transcripts, the exact mechanisms at work could not be elucidated yet. More recently it was shown, that RNase P is required for the formation of the Pol III initiation complex (Serruya et al. 2015). Further studies will be needed to fully elucidate the multiple roles of RNase P.

1.6. MRP RNA

In eukaryotes the RNase P has a close cousin, the RNase mitochondrial RNA processing (RMRP). Originally it was discovered in mouse mitochondria, where it plays a role in DNA replication (Chang & Clayton 1987). The majority of RMRP however, is found in the nucleus, where it intervenes in the processing of pre-rRNA in both yeast and humans (Schmitt & Clayton 1993; Goldfarb & Cech 2017). Furthermore, it was found that RMRP plays a role in cell cycle progression, by assisting in the degradation of cyclin B2 mRNA (Gill et al. 2004). However, the probably most fascinating role of the MRP RNA has been discovered by Maida et al. (2009). They were able to show that in human cells the MRP RNA associates with the telomerase to form a ribonucleoprotein complex, which serves as an RNA dependent RNA polymerase (RdRP). This RdRP can produce double stranded RNAs which then are cleaved into small interfering RNAs.

1.7. 7SL RNA

The 7SL RNA is an abundant cytoplasmic RNA that forms the backbone of the signal recognition particle (SRP). This particle is a cytoplasmic ribonucleoprotein complex that will guide nascent proteins to the endoplasmatic reticulum. In humans the SRP is formed of the 7SL RNA and six proteins: SRP9, SRP14, SRP19, SRP54, and the SRP68/72 heterodimer (Hu et al. 2012). Phylogenetic studies showed that the 7SL RNA is at the origin of several short interspersed elements (SINEs) in the human genome (Ullu & Tschudi 1984; Kriegs et al. 2007)

1.8. Vault RNAs

In the human genome three genes code for vault RNAs (HVG1-3) with a size of 88 – 141 bases (Van Zon et al. 2001). These RNAs bind to the vault particle, which in mammals consists of three proteins: the major vault protein (MVP), the telomerase-associated protein (TEP1) and the vault poly(AD-P) ribose polymerase (vPARP).

Vault particles are abundantly expressed and with a size of 12.9 MDa they are even bigger than ribosomes. They were first discovered in 1986 (Kedersha & Rome 1986) and since then they have been identified in a variety of species ranging from protozoans to mammals, with some notable exceptions such as *Saccharomyces cerevisiae, Caenorhabditis elegans* or *Drosophila melanogaster* (Berger et al. 2009).

Maybe it is due to the lack of vault particles in some of biologists' favorite model organisms that knowledge on the role of vaults is still sparse. Vault particles have been linked to cellular trafficking, signal transmission, immune response and drug resistance, but no clear picture has emerged so far (Suprenant 2002; Steiner et al. 2006; Berger et al. 2009).

1.9. Y RNAs

There are four Y RNAs in the human genome: hY1, hY3, hY4 and hY5. They are relatively small with roughly 100 base pairs and they all form characteristic stem loop structures. Homologous to Y RNA have been found in prokaryotes, nematodes, insects and several vertebrates (Kowalski & Krude 2015).

Y RNAs are most well known for their interaction with the Ro60 protein, with which they form the Ro ribonucleoprotein (RoRNP). Mounting evidence points to a role of RoRNP in RNA processing and quality control (O'Brien and Wolin, 1994; Chen *et al.*, 2003; Sim and Wolin, 2011; Hall *et al.*, 2013). But Y RNAs also have a role independent of the RoRNP. In an *in vitro* system it has been shown that Y RNAs are essential for DNA replication (Christov et al. 2006). The exact mechanisms at work are still unknown, but a recent study showed that Y RNAs associates with chromatin in concordance with the origin recognition complex (ORC) (Kheir & Krude 2017).

1.10 7SK RNA

After transcription initiation the RNA polymerase II (Pol II) pauses due to the effect of inhibitory factors. Such paused polymerases can be found on most genes (Guo & Price 2013), therefore it is most important that release from the paused state is tightly regulated. One important regulator is the positive transcription elongation factor (p-TEFb), it comprises the cyclin T1 and the cyclin dependent kinase Cdk9. It was shown that 7SK is necessary to inhibit the activity of p-TEFb, but it is not sufficient (Nguyen et al. 2001). Indeed, the inhibition is conveyed *via* the hexamethylene bisacetamide inducible protein (HEXIM). 7SK is necessary to form the HEXIM:p-TEFb complex (Yik et al. 2003). A recent study showed that 7SK mediates the formation of the HEXIM:p-TEFb complex by a structural change from an open to a closed complex (Brogie & Price 2017).

1.11. BC1 and BC200 RNA

First identified in 1987 (Watson & Sutcliffe), the Brain Cytoplasmic RNA 1 (BCYRN1 or BC200 in humans and BC1 in mice) is a 200 nucleotide long RNA, which is found mainly in the cytoplasm. It is highly overexpressed in brain tissue and slightly elevated in testes, ovary and small intestine (Booy et al. 2017).

BC200 has been shown to inhibit translation (Wang et al. 2002; Kondrashov et al. 2005), namely through the binding to the two heterogeneous nuclear ribonucleoproteins (hnRNP) E1 and E2 (Jang et al. 2017). Deregulation of BC200 levels can therefore have severe consequences. A study showed that the brains of Alzheimer patients show a significant upregulation of BC200 (Mus *et al.*, 2007).

High levels of BC200 have also been found in a number of cancers, notably breast, esophagus and lung cancers (Chen et al. 1997). Recently studies showed that BC200 expression is elevated in dividing cells and a knock-out leads to reduced cell viability and tumor growth (Booy et al. 2017; Singh et al. 2016).

1.12. Virus encoded RNAs

Viruses that infect a cell often do not have their own polymerase, but make use of the cells transcription apparatus. The RNA polymerase III is hijacked for the transcription of several such viral RNAs. The most well known are EBER1 and EBER2 genes of the Epstein-Barr virus, as well the virus associated (VA) RNAs from the adenovirus.

The Epstein-Barr virus is part of the herpesvirus-family and even a benign infection will lead to a lifelong latent presence of the virus. Today it is estimated that 95% of the adult population are carrier of the virus (Moss et al. 2014). The EBER1 and EBER2 RNA are both about 170 nucleotides long and typically found during latency of the virus (Young & Rickinson 2004).

EBER1 was found to interact with a number of proteins from the host cell, including chaperone LA/SSB (Lerner et al. 1981), the ribosomal protein L22 (Fok *et al.*, 2006), as well as three Heterogeneous nuclear ribonucleoproteins (hnRNPs): A1, A2/ B1, and D/AUF1 (Lee *et al.*, 2012). But many mysteries about the function of EBER1 remain to be elucidated. Even less is known of EBER2, which has one confirmed interaction partner, the protein LA (Lerner et al. 1981), but the function of EBER2 remains elusive.

In case of the adenovirus the RNA polymerase III transcribes different VA RNAs. All adenoviruses have at least one VA RNA, which can vary considerably depending on the serotype. These VA RNAs are synthesized during the late stages of viral infection (Reich *et al.*, 1966; Söderlund *et al.*, 1976) and counteract the host cell defense mechanisms. It has been known for long time that VA-I interacts with the human double-stranded RNA-activated protein kinase R (PKR) thereby enabling viral protein

translation (Kitajewski et al. 1986). Only more recently it was found that VA RNAs can also interfere with DICER there by limiting RNA interference (Andersson et al. 2005).

1.13. SINE and ALU

The human genome encodes around 20 000 genes, but even if this number is impressive, genes make up for less than 2% of total DNA (Lander et al. 2001; Ezkurdia et al. 2014). A much broader group is formed by the repetitive and transposable elements (TE), they make up for roughly 45% of the genome (Chen & Carmichael 2008). The TEs can be divided into two main classes: DNA transposons, that can move independently in the genome and retrotransposons, which need to be transcribed into RNA, transported to the cytoplasm and retro-transcribed into DNA before they can reintegrate the genome. Depending upon their insertions sites, TEs can disrupt the genome in a potentially catastrophic way. In yeast it was shown that an interaction between the Pol III subunit AC40 and the retrotransposon Ty1, leads to direct integration in front of Pol III transcribed genes. Thereby preventing integration in a harmful site, e.g. inside a protein coding gene (Bridier-Nahmias et al. 2015).

While most of the retrotransposons are transcribed by the RNA polymerase II, one group is transcribed by the RNA polymerase III (Kramerov & Vassetzky 2011). This group – the short interspersed elements (SINEs) – is made up of small RNAs that are between 85 to 500 base pairs long (Elbarbary *et al.*, 2016). All SINEs are derived from Pol III transcripts, namely tRNAs, 7SL RNA and 5S rRNAs (Kapitonov & Jurka 2003; Wicker et al. 2007; Kramerov & Vassetzky 2011).

While transposable elements were long timed termed “junk-DNA” or “selfish-DNA”, it is now known that SINEs are accelerators of evolution and fulfill many new functions in the genome. SINEs have been reported to regulate gene expression, by serving as promoters, enhancers, silencers or insulators. Furthermore, they are implicated in alternative splicing, polyadenylation and they can act as trans-factors of transcription, translation or mRNA stability (Makałowski, 2000; Ponicsan *et al.*, 2010; Gong and Maquat, 2011; Lunyak and Atallah, 2011; Elbarbary *et al.*, 2016).

The most well studied class of SINEs are Alu elements. They originated from a truncated version of the 7SL RNA, sometime before the primate/rodent evolutionary divergence (Kriegs et al.,2007). Given their relatively short existence, they are incredibly abundant. Most Alu elements are found in gene rich regions, where they serve a number of functions such transcription regulation or alternative splicing (Chen & Yang 2017). For example it was shown that Alu elements can repress Pol II transcription upon heat shock (Mariner et al. 2008). Alu elements are so abundant in gene rich regions, it is estimated that an average pre-mRNA contains 16 Alu elements (Chen & Carmichael 2008). The progression of Alu elements still continues today with approximately one Alu insert per 20

births in humans. Thus the genomes of any two individuals possess about 800 Alu polymorphisms (Deininger et al. 2011).

1.14. Other transcripts

The above described RNAs represent the most well studied Pol III transcripts, but the list is not exhaustive. A number of studies have tried to identify new Pol III targets using chromatin immunoprecipitation sequencing (ChIP-seq) (Canella *et al.*, 2010; Moqtaderi *et al.*, 2010; Oler *et al.*, 2010). Indeed, the studies were able to identify numerous new Pol III binding sites, but it remains to be seen to what extent these binding sites equal new Pol III transcripts. The search for new Pol III transcripts has only revealed a handful of hitherto unknown targets (Carnevali et al. 2017; Conti et al. 2015; Dieci et al. 2013). But even if the transcription by Pol III was confirmed for these genes, nothing is known of their function.

One exception is the discovery of the non-coding RNA *nc886* (Lee et al. 2011; Lee 2015). It is transcribed by Pol III and seems to be related to the vault RNAs. *Nc886* has been shown to inactivate the Protein Kinase R (PKR), has reduced expression levels in some cancers and a knockdown of *nc886* is sufficient for PKR activation. This led the authors to propose a model in which *nc886* functions as a tumor sensor. During tumorigenesis *nc886* becomes inactivated *via* island methylation. The loss of *nc886* leads to an activation of PKR, which will drive cells into apoptosis.

Undoubtedly the Pol III transcriptome holds many more mysteries and future studies will need to identify new transcripts and also their respective roles.

2. The Pol III transcription apparatus

Before a polymerase can transcribe a gene, the polymerase needs to be recruited to the transcription start site (TSS). This is done at specific locations near the transcription start sites called promoters. The RNA polymerase III has a very limited set of promoters, which are traditionally divided into three classes accordingly named promoters type I, II and III. There also exist a number of genes that mix elements from different promoter types. A non-exhaustive overview of gene classes transcribed by each promoter is given in table 1. The promoters type I and type II are located downstream of the

Table 1: Overview of promoters found in different genes transcribed by the human Pol III

Promoter	RNAs transcribed
Type I	5S rRNA
Type II	tRNA and adenoviral RNA (VA RNA)
Type III	U6 snRNA, RNAse P RNA, RNAse MRP RNA, Y RNA and 7SK RNA
Mixed	7SL RNA, vault RNAs, BC1 and BC200 RNA, EBER RNA

TSS, whereas the promoter type III has upstream regulatory elements (figure 2).

Each promoter has a set of transcription factors (TFs) that are active in recruiting the RNA polymerase III. Some of these transcription factors are common between different promoters, others are unique to just one type of promoter. Besides the transcription factors there are other mechanisms that influence transcription, such as epigenetic markers and nucleosome positioning.

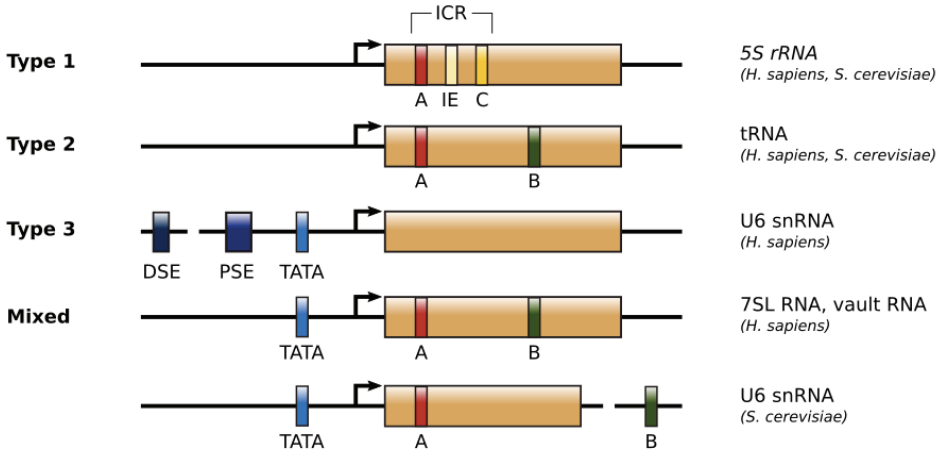

Figure 2: Promoter types of the RNA polymerase III. The type 1 promoter contains an A-box (A), an internal element (IE) and a C-box (C). Together they form the internal control region (ICR). The transcription start site is marked by an arrow. The type 2 promoter has an A- and a B-box, which are both downstream from the TSS start site. Only upstream regulatory elements are found in the type 3 promoter. It contains a TATA-box, a proximal and a distal sequence element (PSE and DSE respectively). Some transcripts, like the human 7SL and vault RNA or the yeast U6 snRNA, combine elements from different promoter types.

2.1. Promoter type I

The type I promoter was first discovered in *Xenopus laevis* (Bogenhagen *et al.*, 1980; Sakonju *et al.*, 1980). It possesses three distinct genetic elements, all located downstream of the transcription start site: the A-box (+50 / +60), the intermediate element (IE) (+67 / +72) and the C-box (+80 / +97). The three elements together form the internal control region (ICR), which is conserved in different species with some exceptions. Notably it was found that in *Saccharomyces cerevisiae* only the C-box is required for transcription (Challice & Segall 1989).

The type I promoter is only found in the genes coding for 5S rRNA. During transcription initiation the transcription factor TFIIIA will bind to the ICR, which allows the transcription factor TFIIIC to be recruited. TFIIIC binds to the TFIIIA:DNA complex and recruits in turn TFIIIB. It is TFIIIB that ultimately interacts with the RNA polymerase III to allow for transcription initiation (reviewed in Schramm and Hernandez, 2002).

2.2. Promoter type II

One of the most abundant RNAs transcribed by Pol III are tRNAs. They possess a type II promoter, which, like the type I promoter, is located downstream of the TSS (Galli *et al.*, 1981; Hofstetter *et al.*, 1981; Sharp *et al.*, 1981). Type II promoters consist of an A-box (+8 / +19) and a B-box (+45 / +62). While the distance between the A-box and the TSS is fix, the B-box can vary in its distance to the A-box to accommodate for introns (reviewed in Dieci *et al.*, 2007).

The A-boxes of the type I and II promoter show structural similarities and are interchangeable in *Xenopus laevis* (Ciliberto et al. 1983). Even though the two A-boxes are similar, they do not serve the same function in the two promoter types. In the type II promoter the A- and B-box do not need TFIIIA to recruit TFIIIC, but rather bind directly to TFIIIC. In turn TFIIIC will then again interact with TFIIIB to recruit RNA polymerase III (reviewed in Schramm and Hernandez, 2002).

2.3. Promoter type III

While the type I and type II promoters have only gene internal elements, the type III promoter is comprised solely of gene external elements. It is composed of at least a proximal sequence element (PSE) at around -50 and a TATA-box at position -30. In vertebrates this minimal promoter is often associated with a distal sequence element (DSE) located approximately 200 bases upstream of the TSS enhances transcription from the PSE:TATA core promoter (Ullu and Weiner, 1985; Murphy, *et al.*, 1986; Bark *et al.*, 1987; Krüger and Benecke, 1987; Das *et al.*, 1988).

Interestingly the PSE is also a core element of Pol II promoters, it is the presence of the TATA-box that decides whether a promoter is used by Pol II or Pol III (Mattaj et al. 1988; Lobo & Hernandez 1989). For example, the snRNAs U1, U2, U4 and U5 are transcribed by RNA polymerase II, whereas U6 is transcribed by Pol III. The only difference between the promoter of U6 compared to the other U genes is the presence of a TATA-box (Kunkel and Pederson, 1988; reviewed in Jawdekar and Henry, 2008). Mutating the TATA-box will induce Pol II transcription of the U6 gene (Lobo & Hernandez 1989). Inversely the addition of a TATA-box to a U2 gene, normally transcribed by Pol II, will lead to transcription by Pol III (Mattaj et al. 1988).

Recruitment of Pol III to type III promoters starts with the binding of the snRNA activating protein complex (SNAPc), also referred to as the PSE-binding protein (PBP) or the PSE transcription factor (PTF). This complex will in turn recruit TFIIIB, which will contact the RNA polymerase III (reviewed in Schramm and Hernandez, 2002). *In vitro* it was shown that TFIIIB alone is sufficient for Pol III recruitment (Teichmann et al. 1997).

2.4. Mixed promoters

While the promoters of type I-III make up for a large number of genes transcribed by Pol III, there are some exceptions. Vault and 7SL RNAs contain internal A- and B-boxes like the type II promoter, but also possess an external TATA-box (Stadler et al. 2009; Englert et al. 2004). Another gene coding for a selenocysteine tRNA in *Xenopus laevis* has an internal B-box, but no A-box. Instead it possesses external DSE, a SNAPc binding site and a TATA-box (Carbon & Krol 1991). CHiP-seq studies revealed that vault and 7SL RNAs were associated with proteins linking them to TFIIIC, as in the case of a type II promoter. The $tRNA^{Sec}$ on the other hand is more likely associated with factors from the type III promoter class (Canella *et al.*, 2010; Moqtaderi *et al.*, 2010).

Other examples of mixed promoters exist (Brow & Guthrie 1990; Martignetti & Brosius 1993; Gogolevskaya & Kramerov 2010) and along with the discovery of new Pol III transcripts, further mixed promoter genes might be identified. It remains to be seen if these kind of promoters remain the exceptions or if they are more abundant than it seems today.

2.5. Transcription termination

Even if the different transcripts synthesized by Pol III possess a variety of promoters, they all have the same transcription termination signal. A simple stretch of d(T)s is sufficient for Pol III transcription termination (Bogenhagen *et al.*, 1981; Cozzarelli *et al.*, 1983; Arimbasseri *et al.*, 2013). The number of T repeats depends on the species. While 4Ts are enough in many vertebrates including humans, it needs 5 and 6 Ts respectively in *Saccharomyces pombe* and *cerevisiae* (Bogenhagen et al. 1981; Cozzarelli et al. 1983; Allison & Hall 1985; Hamada et al. 2000).

However, it has been observed that the RNA polymerase III can read through longer T stretches. The most striking example is the one of the SNR52 transcription unit in *Saccharomyces cerevisiae.* A T_6 stretch is located between the A- and B-box. Even though this is the classic termination signal for Pol III in *S. cerevisiae*, the polymerase can read through the signal without a problem (Braglia *et al.*, 2005). Other studies confirm that read through events are far from being a rare exception (Matsuzaki *et al.*, 1994; Nielsen *et al.*, 2013; Rijal and Maraia, 2016). Therefore, other factors must play a role in transcription termination.

Early studies already suggested that the sequence context around upstream and downstream of the T stretch is important for the efficiency of termination (Bogenhagen et al. 1981; Cozzarelli et al. 1983). Since then a number of studies have found more evidence for a dependence on sequence context of Pol III, but no general rule could be identified (Mazabraud *et al.*, 1987; Chu *et al.,* 1997; Gunnery *et al.*, 1999; Braglia *et al.*, 2005).

Besides sequence context there may also be trans-acting factors that regulate transcription termination. One of the first candidates to be identified was the La protein (Stefano 1984). While a number of studies confirm an involvement of La in transcription termination (Gottlieb and Steitz, 1989; Maraia *et al.*, 1994; Goodier and Maraia, 1998; Maraia and Lamichhane, 2011), others contest such a role (Lin-Marq and Clarkson, 1998; Schramm and Hernandez, 2002; Hu *et al.*, 2003). Other reports on trans-acting elements include the transcription factor TFIIIC (Wang & Roeder 1998) or nuclear factor 1 (NF1) (Wang et al. 2000). Clearly more research is needed to unravel all the mechanisms of Pol III transcription termination.

2.6. Transcription factor III A (TFIIIA)

The first eukaryotic transcription factor to be purified was the Pol III transcription factor III A (TFIIIA), which was purified form *Xenopus laevis* (Engelke et al. 1980). It was also the first TF for which a corresponding cDNA was isolated (Ginsberg *et al.*, 1984). TFIIIA is essential for cell survival as it binds to the type I promoter of the 5S rRNA and thus enables transcription. However, it seems as if this is the only essential function of TFIIIA. For it was shown that yeast strains that had engineered 5S rRNA under type II promoter control and that were depleted of TFIIIA, were viable (Camier *et al.*, 1995).
TFIIIA is poorly conserved among species (Layat *et al.*, 2013). Between humans and *Xenopus laevis* the sequence identity of TFIIIA is 63% on a nucleotide level and only 58% for amino acids, whereas it is 94% for TFIIIC (Arakawa *et al.*, 1995; reviewed in Schramm and Hernandez, 2002). While the sequence of TFIIIA has not been conserved during evolution, its structure remained remarkably unchanged. All TFIIIAs have nine consecutive zinc fingers (Miller *et al.*, 1985), the only exception to the rule is *Saccharomyces pombe* which has ten (Schulman & Setzer 2002). The TFIIIA zinc finger protein has since become the archetype of the C_2H_2 zinc finger class. The zinc fingers bind with high affinity to the ICR of the 5S rRNA gene and once TFIIIA is firmly fixed on the DNA it will bind to TFIIIC, which together with TFIIIB recruits Pol III (figure 3) (Bieker *et al.*, 1985; Kassavetis *et al.*, 1990).

2.7. Transcription factor III C (TFIIIC)

In yeast and in humans TFIIIC is needed for the transcription initiation from promoter types I and II. In case of the type I promoter TFIIIC will bind to TFIIIA, which is attached to the DNA. In case of the type II promoter however, TFIIIC can bind directly to the DNA. In yeast TFIIIC is composed of six subunits Tcf1 (τ95), Tcf3 (τ138), Tcf4 (τ131 or PCF1), Tcf6 (τ91), Tcf7 (τ55) and Tcf8 (τ60) (Swanson et al. 1991; Lefebvre et al. 1992; Marck et al. 1993; Arrebola et al. 1998; Manaud et al. 1998; Deprez et al. 1999).

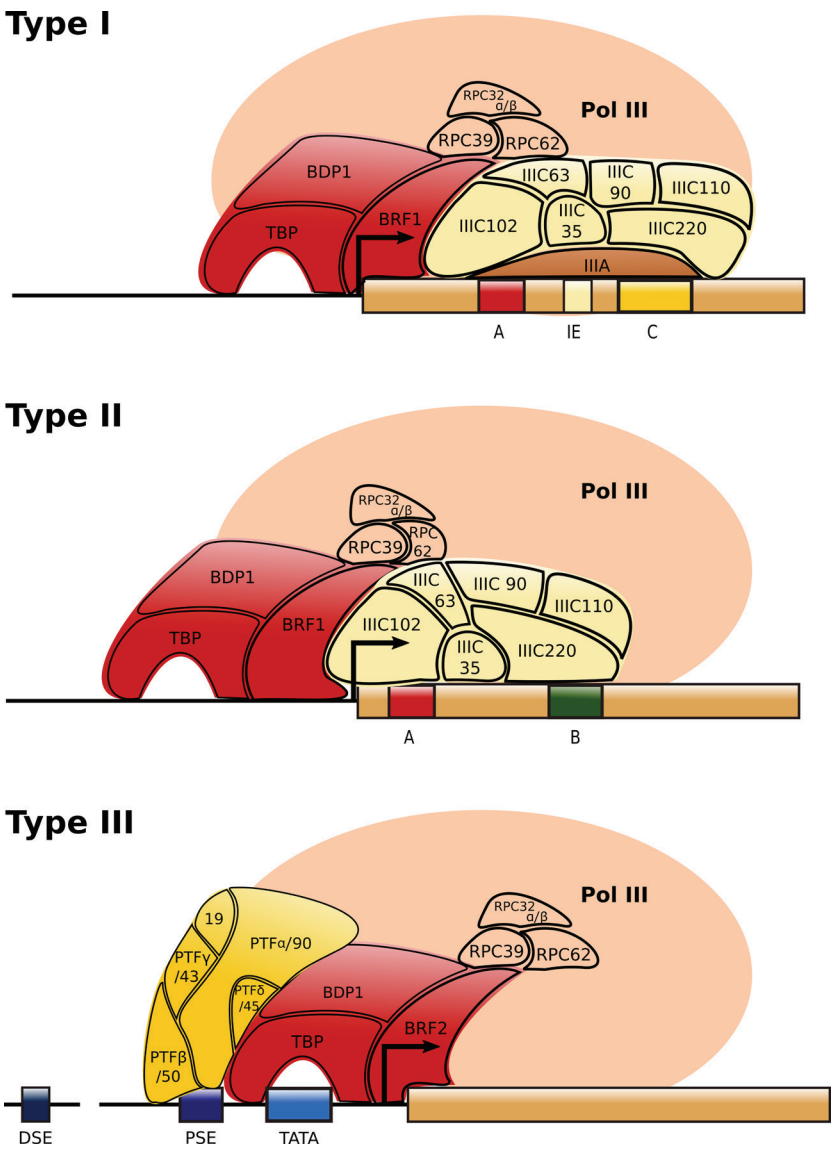

Figure 3: Promoters and transcription factors of the RNA polymerase III (Pol III). The type 1 promoter consists of an A- and C-Box which are bound by transcription factor III A (TFIIIA) (brown). The multi-protein transcription factor III C (TFIIIC) (light yellow) binds TFIIIA and contacts the transcription factor III B (TFIIIB) (red). It is TFIIIB that recruits Pol III (orange) to the promoter. In case of the type 2 promoter the roles are the same, but TFIIIC does not need TFIIIA to bind to the A- and B-box. The type 3 promoter differs from the first two as it has only external promoter elements. The transcription factors SNAPc (dark yellow) is indicated by its PTF names and the molecular weight of the respective SNAPc subunits. SNAPc will bind to the proximal sequence element (PSE).

The six subunits are grouped into two domains, τA et τB, that are connected *via* a flexible linker (Ruet et al. 1984; Marzouki et al. 1986; Schultz et al. 1989). τA and τB will bind to the A- and B-box respectively, the flexible linker allows for variable distances between the two DNA domains (Baker et al. 1987). Electron microscopy revealed that the τA domain is made up of three subunits: Tcf1 (τ 95), Tcf4 (τ131) and Tcf7 (τ55) (Schultz et al. 1989). The other three subunits Tcf3 (τ138), Tcf6 (τ91) and Tcf8 (τ60) form the τB domain (Deprez et al. 1999; Arrebola et al. 1998; Lefebvre et al. 1992). These data were confirmed by a reconstitution of the domains τA and τB using the baculovirus expression

system (Ducrot et al. 2006). The reconstituted TFIIIC was functionally equivalent to a TFIIIC purified from yeast. This proved that six subunits are sufficient for an active TFIIIC complex.

The subunits Tcf1 (τ95) and Tcf3 (τ138) will contact the DNA at the A- and B-box respectively (Gabrielsen et al. 1989; Bartholomew et al. 1990). Tcf4 (τ131) is the only subunit that protrudes upstream of the TSS into a region occupied by TFIIIB (Braun et al. 1992). *Via* coimmunoprecipitation and two hybrid experiments it was shown that Tcf4 (τ131) interacts with Bdp1 and Brf1, two subunits of TFIIIB (Chaussivert et al. 1995; Rüth et al. 1996; Dumay-Odelot et al. 2002; Male et al. 2015). Furthermore, an *in vitro* and *in vivo* interaction was found between Tcf4 (τ131) and the Pol III subunit Rbp12 (Dumay et al. 1999). The third subunit of TFIIIB, TBP, will bind to the C-terminal domain of Tfc8 (τ60) (Deprez et al. 1999).

In humans it was shown that TFIIIC can be separated into two fractions named TFIIIC1 and TFIIIC2 (Dean and Berk, 1987; Yoshinaga *et al.*, 1987). Both fractions are necessary and sufficient to reconstitute TFIIIC activity in *in vitro* transcription experiments (Wang & Roeder 1998). TFIIIC2 is the homolog of yeast TFIIIC and correspondingly it is composed of six subunits: TFIIIC220, TFIIIC110, TFIIIC102, TFIIIC90, TFIIIC63, TFIIIC35 (L'Etoile *et al.*, 1994; Lagna *et al.*, 1994; Sinn *et al.*, 1995; Hsieh, Kundu, *et al.*, 1999; Hsieh *et al.*, 1999; Dumay-Odelot *et al.*, 2007). The yeast homolog and function of each subunit is given in table 2.

The second fraction of human TFIIIC, TFIIIC1, seems to have no homolog in yeast. In humans it is required for transcription from all three promoter types (Yoon et al. 1995), but its exact function remains somewhat elusive. The TFIIIC1 fraction contains one TFIIIC complex with four subunits of 70, 50, 45 and 40 kDa (Wang & Roeder 1998). While this complex has no strong DNA-binding activity of its own, it will help stabilize the TFIIIC2:DNA interaction in type II promoters (Yoshinaga *et al.*, 1987). In the case of type I promoters TFIIIC1 enhances the binding of TFIIIA to the ICR and it improves the

Table 2: Homologies between yeast and human TFIIIC subunits and their respective function.

S. cerevisiae	***Homo sapiens***	**Function**
Tcf1 (τ95)	TFIIIC63	Binds to the A-box of type II promoters, can form a subcomplex with Tcf7 / TFIIIC35, binds to hBRF1, hTBP, TFIIIC102 and RPC62
Tcf3 (τ138)	TFIIIC220	Binds to the B-box together with Tcf6 / TFIIIC110, has an intrinsic HAT activity
Tcf4 (τ131 or PCF1)	TFIIIC102	Binds to hBRF1, hTBP and TFIIIC63
Tcf6 (τ91)	TFIIIC110	Binds to the B-box together with Tcf3 / TFIIIC220, has an intrinsic HAT activity
Tcf7 (τ55)	TFIIIC35	Binds to the A-box. Can form a subcomplex with Tcf1 / TFIIIC63
Tcf8 (τ60)	TFIIIC90	Connects τA and τB, binds to yTBP, TFIIIC220, 110, 63, hBRF1, RPC62 and RPC39, has an intrinsic HAT activity

stability of the PBP/PTF/SNAPc complex in type III promoters (Oettel et al. 1997). *In vitro* transcription experiments suggest that the TFIIIC1 complex is functionally related to BDP1 of the TFIIIB (Weser et al. 2004).

Besides the TFIIIC1 complex, the TFIIIC1 fraction contains also a factor that enhances human U6 transcription and which was accordingly named TFIIICU (Oettel *et al.*, 1998). Finally the TFIIIC1 fraction contains nuclear factor 1 (NF1) peptides, that play a role in transcription termination (Wang et al. 2000). This shows that the role of human TFIIIC extends beyond that of a simple recruitment factor for TFIIIB.

Among the different functions of TFIIIC is also the facilitation of Pol III reinitiation. Quick reloading of the polymerase to the transcription start site is of high importance to maintain high efficiency of transcription (Dieci & Sentenac 1996). For short genes (≤ 100 bp) TFIIIB is sufficient to reinitiate transcription, probably because the polymerase III is still in the vicinity of the TSS. On longer genes (> 300 bp) however, TFIIIC is necessary for high reinitiation rates (Ferrari et al. 2004). As TFIIIC is bound to the DNA downstream of the TSS, it stays in contact with Pol III after transcription initiation and helps to reestablish contact with TFIIIB, which remains bound to the promoter. TFIIIC will facilitate reinitiation even if the B-box is placed as far as 600 bp downstream of the TSS (Ferrari et al. 2004).

The fact that TFIIIC can bring the polymerase back in contact with TFIIIB after transcription termination, brings up the question of what happens to TFIIIC during transcription? Does it stay bound to the DNA or is it replaced by the advancing Pol III? Different *in vitro* studies came to opposing conclusions. Observations made by Ruet *et al.* (1984) seem to suggest that TFIIIC remains bound to the DNA during Pol III transcription. Whereas Bardeleben *et al.* (1994) found that TFIIIC is readily displaced by the advancing polymerase. A ChIP study identified a low but persistent occupancy of TFIIIC on transcribed genes (Soragni & Kassavetis 2008). Recently a study analyzed nascent transcripts attached to the Pol III and found that the polymerase is distributed very unevenly along the transcripts. Two major occupational peaks were identified that corresponded to the A- and B-box. The authors hypothesize that TFIIIC needs to be displaced from the DNA before the polymerase can advance and the two occupational peaks correspond to pausing Pol III awaiting the displacement of TFIIIC (Leśniewska & Boguta 2017).

Three subunits of TFIIIC, TFIIIC220, TFIIIC110 and TFIIIC90, possess a histone acetyltransferase (HAT) activity (reviewed in Schramm and Hernandez, 2002). The acetylation of histones is often linked to the opening of the chromatin, which allows for transcriptional activation and indeed TFIIIC was shown to release transcriptional repression due to chromatin remodeling (Kundu *et al.*, 1999). More recently, using chromatin immunoprecipitation, a study identified eight loci that were occupied by TFIIIC, but not by the rest of the Pol III machinery (Moqtaderi & Struhl 2004). These loci, termed extra

TFIIIC loci (ETC), were preserved in different *Saccharomyces* species which led the authors to believe that they served a conserved function. Since then a number of studies have confirmed a role of TFIIIC in genome organization (Moqtaderi et al. 2010; Noma et al. 2006; Valenzuela et al. 2009; Donze 2012; Van Bortle & Corces 2012).

2.8. Transcription factor III B (TFIIIB)

In Pol III there are three different promoter types, each with its own transcription initiation mechanism. The final step for all three mechanisms is the recruitment of Pol III *via* TFIIIIB. During *in vitro* experiments it was shown that TFIIIB is capable to correctly position Pol III on both type I and II promoters and to initiate transcription, even in the absence of TFIIIC (Kassavetis et al. 1990). Therefore, TFIIIB is the key transcription factor for all Pol III transcripts.

The yeast TFIIIB is composed of three subunits: the TATA-binding protein (TBP) (Hahn et al. 1989; Horikoshi et al. 1989), the TFIIIB related factor 1 (Brf1) (Buratowski et al. 1992; Colbert & Hahn 1992) and B double prime 1 (Bdp1) (Kassavetis et al. 1995; Rüth et al. 1996). In drosophila TBP is replaced by an ortholog named TBP-related factor 1 (TRF1) (Takada et al. 2000).

While in yeast and drosophila one single type of TFIIIB is sufficient for transcription of all Pol III transcripts (Joazeiro *et al.*, 1994), two types of TFIIIB are needed in higher eukaryotes (Lobo et al. 1992). One form, TFIIIBα, containing TBP, BDP1 and BRF2 is used in transcription from external type III promoters. Another form, TFIIIBβ, comprised of TBP, BDP1 and BRF1 recruits Pol III to the internal type I and II promoters. (Teichmann & Seifart 1995; Schramm et al. 2000).

TBP is a subunit of a common transcription factor for RNA polymerases I, II and III. It binds specifically to a TATAAAA motif in the minor grove of the DNA. The crystal structure of TBP shows that it sits on the DNA and bends it dramatically (figure 4A). This widens the minor groove and brings transcription factors bound on the promoter in closer proximity (Kim *et al.*, 1993; Nikolov *et al.*, 1996; reviewed in Orphanides *et al.*, 1996)

In yeast TBP can initiate correctly the assembly of TFIIIB on the U6 gene, even in absence of TFIIIC (Margottin *et al.*, 1991; Joazeiro *et al.*, 1994). It will not only bind the TATA-box containing type III promoters, but also the TATA-less type I and II promoters, but in this case the presence of TFIIIC is required (White & Jackson 1992). In *E.coli* produced recombinant proteins TBP, Brf1 and Bdp1 could successfully be assembled into a fully functional recombinant TFIIIB (Kassavetis et al. 1995). This work allowed to identify the different assembly steps of TFIIIB. Brf1, TBP and lastly Bdp1 are successively incorporated into the TFIIC:DNA complex (Kassavetis et al. 1991; Kassavetis et al. 1992; Huet & Sentenac 1992).

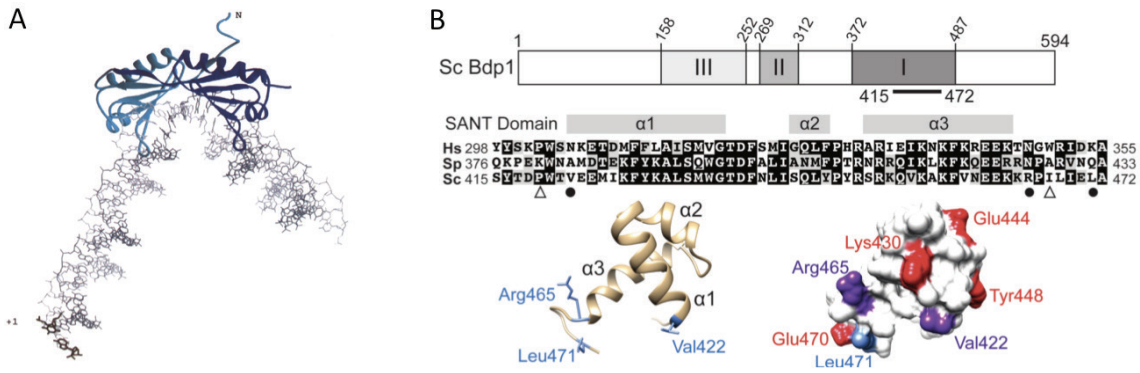

Figure 4: Structure of the TBP:DNA interaction and the Bdp SANT domain. (A) The TFIIIB subunit TBP (blue ribbons) sits on the DNA and bends it, thereby widening the minor groove. (B) Essential regions I, II and III of the *S. cerevisiae* Bdp1 SANT domain. Sequence alignment is given with the Bdp1 homologs of *H. sapiens, S. pombe* and *S. cerevisiae*. Identical and similar residues are shaded in black and grey respectively. On the bottom are shown the structural models of the SANT domain in backbone trace (left) and the molecular surface (right). The blue side chains of the backbone trace model are involved in the binding to Brf1. The molecular surface model shows the residues involved in Brf1 interaction based on the method of detection: NMR analysis (red), NMR analysis and BPA cross-linking (purple) and only BPA cross-linking (blue).

Images from Orphanides *et al.,*1995 and Hu *et al.,* 2015

Brf1 is a key component of TFIIIB. It binds both TFIIIC (subunit Tfc4) (Khoo *et al.*, 1994; Chaussivert *et al.*, 1995) and the RNA polymerase III (subunits C34 and C17) (Werner et al. 1993; Ferri et al. 2000). Bdp1 is the largest subunit of TFIIIB. It possess a SANT motif (found in SWI-SNF, ADA2, N-CoR and TFIIIB), which interacts with DNA and Brf1 (Aasland *et al.*, 1996; Kassavetis *et al.*, 2006) (figure 4B). BRF1 and BRF2 associate tightly with TBP and can be co-purified, whereas BDP1 is only loosely connected to the complex in absence of DNA (Kassavetis et al. 1991). In presence of DNA however, Bdp1 will contribute to form a very stable TFIIIB:DNA connection, which has recently been confirmed by structural studies (Gouge et al. 2017). The bond to the DNA is so strong that TFIIIB can act as a roadblock to advancing RNA polymerase II (Roy et al. 2016). Bdp1 also plays an important role in the opening of the transcription bubble, as defective Bdp1 mutants were still able to recruit Pol III, but could be rescued by preopening DNA templates (Kassavetis et al. 2003; Hu et al. 2015).

2.9. snRNA activated protein complex (SNAPc)

Type 3 promoters are characterized by the fact that they have only gene external promoter elements: the proximal and the distal sequence element (PSE and DSE respectively). The PSE is bound by the snRNA activated protein complex (SNAPc), also known as PSE-binding factor (PTF) or PSE-binding protein (PBP) (Waldschmidt *et al.*, 1991; Murphy *et al.*, 1992; Sadowski *et al.*, 1996). SNAPc is composed of five subunits: SNAP190/PTFα (Wong et al. 1998), SNAP50/PTFβ (Henry et al. 1996; Bai et al. 1996), SNAP45/PTFδ (Sadowski et al. 1996; Yoon & Roeder 1996), SNAP43/PTFγ (Henry et al.

1995; Yoon & Roeder 1996) and SNAP19 (Henry et al. 1998). In humans SNAPc serves as a transcription factor to both Pol II and Pol III (Henry et al. 1998).

Once bound to the DNA, SNAPc will contact TFIIIB, which will in turn recruit the RNA polymerase III (reviewed in Schramm and Hernandez, 2002). Recently it was shown that the interaction with TFIIIB, is mediated by the TFIIIB subunit BDP1 (Gouge et al. 2017). Deletion of the N-terminal of Bdp1 will diminish the interactions with SNAPc, while deletions of the C-terminal are tolerated. Deletion of both extremities of Bdp1 abolishes the connection to SNAPc. Also it was shown that Bdp1 binds strongly to SNAPc in absence and in presence of DNA (Gouge et al. 2017). The binding is strongest in the absence of DNA, showing that the two proteins interact directly with one another.

The distal sequence element (DSE) serves as a binding site for the factors OCT-1 and STAF (Carbon *et al.*, 1987; Tanaka *et al.*, 1992; Schaub *et al.*, 1997). Interestingly STAF enhances transcription by both Pol II and Pol III (Schaub et al. 1997). OCT-1 on the other hand is specialized in Pol III transcription. Another closely related protein called OCT-2, also binds to the DSE and enhances Pol II transcription (Tanaka *et al.*, 1992).

2.10. Epigenetic factors

For Pol I transcription it has been reported that epigenetic factors help regulate the expression of rRNAs (McStay & Grummt 2008). Like Pol I transcripts, Pol III transcribed RNAs also exist in numerous copies throughout the genome (Canella *et al.*, 2010). It is therefore probable that epigenetic factors intervene in transcription regulation of Pol III as well. Indeed it has been shown that the methylation of Alu elements was tissue specific (Xie et al. 2009) other studies show a decreased Alu methylation in cancer cells (Daskalos et al. 2009; Richards et al. 2009; Xiang et al. 2010).

The acetylation or methylation of histones is another common epigenetic regulation mechanism (Wang *et al.*, 2008; Zhou *et al.*, 2011). Many of the markers that are associated with active transcription in Pol II are also present in Pol III transcribed genes (Barski et al. 2010; Oler et al. 2010) (figure 5). Therefore, it is likely that factors that can change the acetylation status of histones will influence the transcription of Pol III.

While the acetylation status of the histones around Pol II and Pol III transcribed genes is similar, the nucleosome positioning on the gene is quite different. Pol II genes show a nucleosome depleted region (NDR) at the TSS and they possess nucleosomes within the gene-body (Yuan et al. 2005; Shivaswamy et al. 2008). Pol III genes on the other hand seem to be free from nucleosomes (Moqtaderi et al. 2010). A recent study found that unlike Pol II which has a strong positioning of the first nucleosome after the transcription start site (+1 nucleosome), the +1 nucleosome in Pol III is variable (Helbo et al. 2017). The location of the +1 nucleosome varied among cell populations, cell

types and Pol III promoters, which could indicate that the positioning of the +1 nucleosome may regulate Pol III transcription.

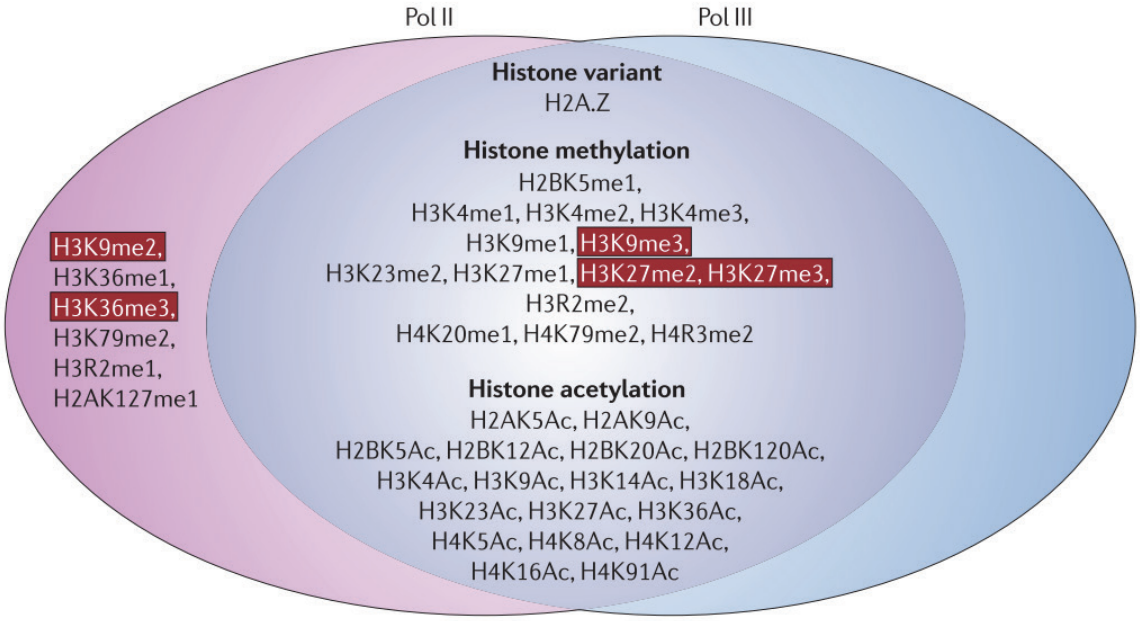

Figure 5: Histone modifications that influence transcription by Pol II and Pol III. Histone modifications affecting Pol II are given in the pink oval, those pertaining to Pol III in the blue oval. Modifications listed in black are known to enhance transcription, those in red box with white lettering do repress transcription.

Image from White, 2011

3. The RNA polymerase III

The three eukaryotic RNA polymerases (RNAP) are all multi protein enzymes of 14, 12 and 17 subunits respectively for Pol I, II and III (figure 6). All three RNAPs possess a core formed of five common proteins and five proteins with significant similarity (Fernández-Tornero et al. 2010). Attached to this 10 subunit core are two proteins that form a protruding stalk, which is involved in transcription initiation (Cramer et al. 2008). These 12 subunits form the RNA polymerase II, which is the most studied of the three polymerases. Varying names exist for the different subunits in yeast and the designations change again from yeast to humans. In the present manuscript a uniform appellation for yeast proteins is used, even if it differs from the original article. An overview of the different terms in use is given in table 3.

Table 3: The RNA polymerases and their subunits across species

Bacteria	Archaea	Eukaryotes						Comment
		Pol I		Pol II		Pol III		
		Yeast	Human	Yeast	Human	Yeast	Human	
Polymerase core								
		Shared subunits						
	Rpo5	Rpb5 *ABC27*	RPB5	Rpb5 *ABC27*	RPB5	Rpb5 *ABC27*	RPB5 *RPABC1, RPB25*	
ω	Rpo6	Rpb6 *ABC23*	RPB6	Rpb6 *ABC23*	RPB6	Rpb6 *ABC23*	RPB6 *RPABC2, RPB14.4*	
	Rpo8	Rpb8 *ABC14.5*	RPB8	Rpb8 *ABC14.5*	RPB8	Rpb8 *ABC14.5*	RPB8 *RPABC3, RPB17*	
	Rpo11	Rpb10 *ABC10β*	RPB10	Rpb10 *ABC10β*	RPB10	Rpb10 *ABC10β*	RPB10 *RPABC5, RPB10β*	
	Rpo12	Rpb12 *ABC10α*	RPB12	Rpb12 *ABC10α*	RPB12	Rpb12 *ABC10α*	RPB12 *RPABC4, RPB10α*	
		Paralogs						
β’	Rpo1	A190 *Rpa190*	RPA190	Rpb1	RPB1	C160	RPC160 *RPC1*	Largest subunit of the polymerase Structural elements: clamp, jaw
β	Rpo2	A135 *Rpa135*	RPA135	Rpb2	RPB2	C128	RPC128 *RPC2*	Second largest subunit Structural elements: clamp, wall
α	Rpo3	AC40 Rpa40	RPA40	Rpb3	RPB3	AC40	RPC40 *RPAC40, RPAC1, RPA5, RPA39*	Shared between Pol I and III Structural elements: back
α	Rpo10	AC19 *Rpa19*	RPA19	Rpb11	RPB11	AC19	RPAC19 RPC16, *RPAC2, RPA9, RPA16*	Shared between Pol I and III Structural elements: back
		A12.2 *Rpa12*	RPA12	Rpb9	RPB9	C11	RPC11 *RPC10*	Structural elements: jaw C-terminal domain resembles TFIIS
Stalk								
	Rpo4	A14	RPA14	Rpb4	RPB4	C17	RPC9 *CRCP*	
	Rpo7	A43	RPA43	Rpb7	RPB7	C25	RPC8 *RPC22.9*	
Heterodimer								
		A49	RPA49			C37	RPC5	Paralog to TFIIFα
		A34.5	RPA34			C53	RPC53 *RPC4*	Paralog to TFIIFβ
Heterotrimer								
						C31	RPC32 α/β *RPC7*	
						C34	RPC39 *RPC6*	Paralog to TFIIEβ
						C82	RPC62 *RPC3*	Paralog to TFIIEα

In Pol I and Pol III a heterodimer is attached to the core, composed of the proteins A34.5 and A49 in Pol I and C53 and C37 in Pol III (Landrieux *et al.*, 2006; Kuhn *et al.*, 2007; Kassavetis *et al.*, 2010). Finally Pol III possesses a further subcomplex formed of three proteins: C31, C34 and C82 (Wang & Roeder 1997) (figure 6). These additional subcomplexes are homologs of Pol II general transcription factors (GTFs). The heterodimer present in both Pol I and III is related to transcription factor IIF (TFIIF), while the heterotrimer of Pol III is related to TFIIE (Cramer et al. 2008; Carter & Drouin 2010). The permanent recruitment of transcription factors allows for a faster assembly and transcription reinitiation. Given that Pol I transcribes only one single gene and Pol III transcribes a limited set of genes, that are important for growth and proliferation, an efficient transcription mechanism is an advantage. The detached system of Pol II on the other hand might be slower, but presents more regulatory options. As Pol II transcribes all protein coding genes, it is essential, that transcription initiation is closely monitored to avoid potentially dangerous mistakes

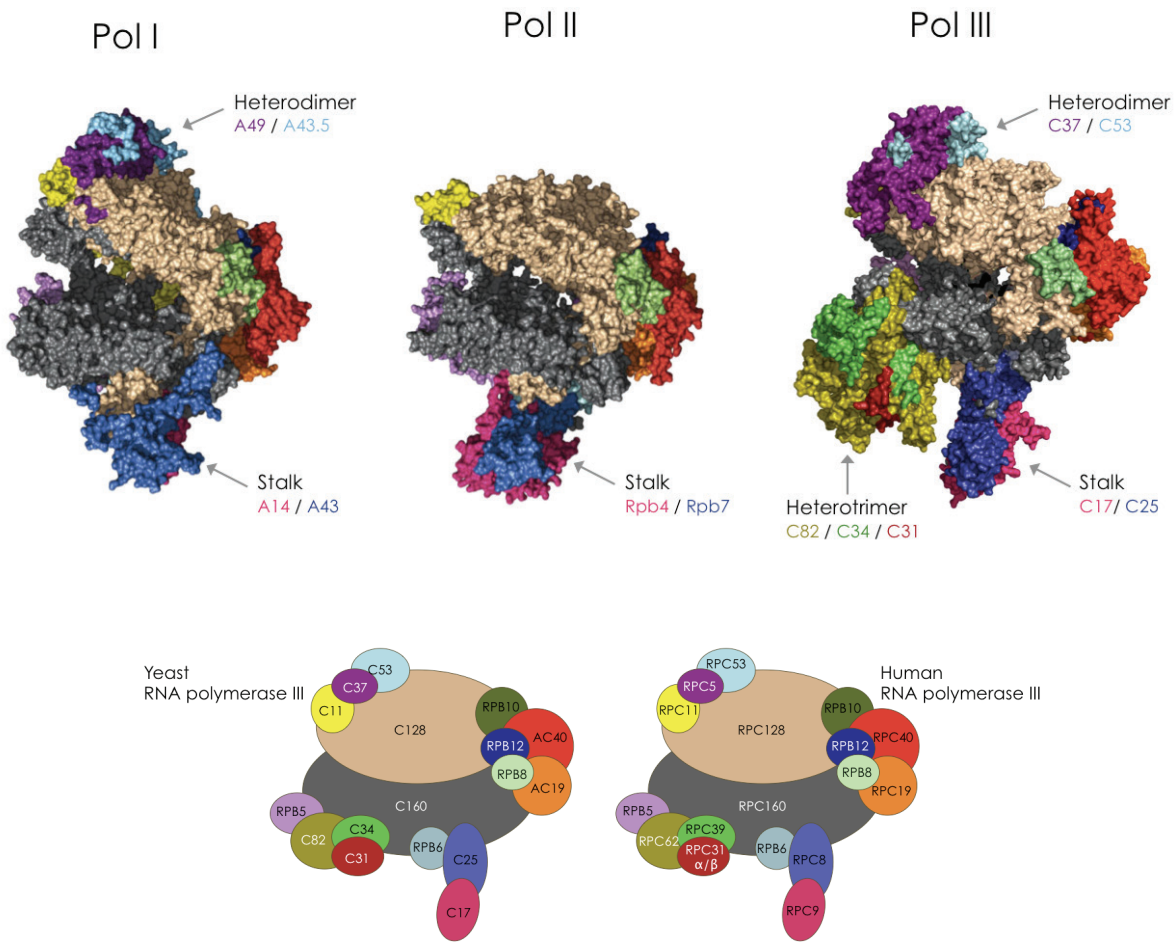

Figure 6: Cryo-EM structure of the eukaryotic RNA polymerase I, II and III. Above is shown a cryo-EM structure of the yeast RNA polymerases I, II and III. Homologous subunits are colored alike. All polymerases have a 10 subunit core, with a two subunit stalk protruding downwards. Pol I and III have an additional heterodimer, here pointing up. Pol III has furthermore a heterotrimer, which is next to the stalk. Below is a schematic view of the yeast and human RNA polymerase III, indicating the names of the different subunits. The subunits that form the heterodimer are colored in blue, those that form the heterotrimer in red.

Cryo-EM structure adapted from Hoffmann *et al.*, 2015

3.1. The core

Like Pol I and Pol II, Pol III has a 10 subunit core. The subunits C160 and C128 are the two largest subunits. They form the center around which the other subunits are grouped. Other subunits of the core are: Rpb5, Rpb6, Rpb8, Rpb10, Rpb12, which are shared by all three polymerases, plus the subunits AC40 and AC19 which are shared by Pol I and III and have homologs in Pol II (see table 3).

The structure of the core of Pol III resembles that of Pol I and II (Cramer et al. 2001; Engel et al. 2013; Fernández-Tornero et al. 2013; Hoffmann et al. 2015). The overall shape resembles a crab claw with a middle cleft that embeds the DNA during transcription. The DNA enters the polymerase through the clamp. It moves along the cleft towards the active site where two pores allow for the entry of nucleotides and the exit of nascent RNA. As the DNA moves on it hits a protein *wall* that directs the DNA at a 90° angle out of the polymerase (Cramer *et al.*, 2000; Hoffmann *et al.*, 2015; Khatter *et al.*, 2017) (figure 7). The total length of the DNA covered by the polymerase stretches from the downstream base pair (+14 bp) in the cleft, through the active site to the upstream base pair (-9 bp) (Hoffmann et al. 2015).

The subunit C11 is particular, as it has two functionally distinct domains. While its N-terminal domains are homologous to the Pol II subunit Rpb9, its C-terminal domain resembles the Pol II elongation and RNA cleavage factor TFIIS (Vannini & Cramer 2012). C11 has been shown to cleave RNA and a lack of this activity hinders transcription termination (Chedin et al. 1998). Furthermore, C11 interacts with the subunit C37 to facilitate transcription termination and efficient reinitiation (Landrieux et al. 2006; Iben et al. 2011).

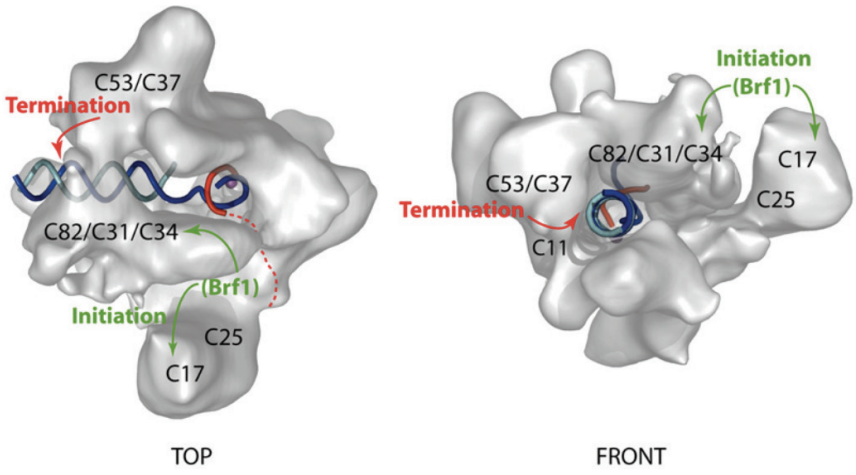

Figure 7: Transcription model for the yeast RNA polymerase III. The DNA coding and non-coding strand are depicted in blue and cyan respectively. In red is shown the nascent RNA strand. The numbers indicate the location of different Pol III subunits as well as the binding sites for TFIIIB subunit Brf1.

Image from Fernández-Tornero *et al.*, 2007

3.2. The stalk

The subunits C17 (RPC9) and C25 (RPC8) form a stalk that protrudes from the core. The two subunits are homologs of A14 and A43 in Pol I, as well as Rpb4 and Rpb7 in Pol II. Compared to its counterparts the stalk of Pol III is larger and extrudes further from the core (Hoffmann et al. 2015; Fernández-Tornero et al. 2007). The stalk is tightly anchored to the core *via* the N- and C-terminal domain of C160. Through a conformational change the stalk can participate in the opening and closing of the clamp. Unbound polymerases show an open clamp conformation that will close during transcription initiation and elongation (Chakraborty et al. 2012).

A study showed that a knock-out of C17 leads to strong defects in tRNA synthesize and cell growth, indicating that C17 is necessary for a correct functioning of Pol III (Ferri et al. 2000). Furthermore, the study showed that C17 can bind to the TFIIIB subunit Brf1, thereby helping with transcription initiation. Another study was able to show that the two subunits C17-C25 have a high affinity for single stranded RNA (Jasiak et al. 2006). This could indicate that the stalk guides the nascent RNA. A theory that is supported by structural models, in which the newly synthesized RNA exits from the core channel next to the stalk (Fernández-Tornero et al. 2007).

The difference between open and closed conformation in Pol III is less pronounced than in other polymerases, leaving a smaller cleft open even in the unbound state (Hoffmann et al. 2015) (figure 8). However, as the stalk moves into the closed clamp position, it pushes the heterotrimer, C82-C34-C31, closer to the DNA in the cleft. The heterotrimer is a unique feature of the Pol III and its close proximity to the DNA could facilitate promoter opening and elongation (Hoffmann et al. 2015).

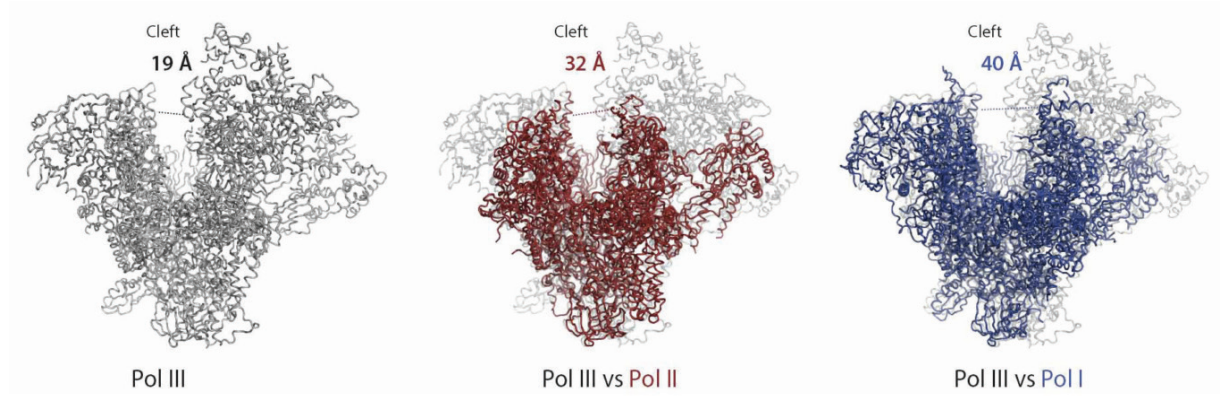

Figure 8: Comparison of the closed clamp formation in Pol III versus Pol II and Pol III. Cryo-EM structure of the closed clamp formations of Pol I, II and III. Pol III is illustrated in grey, Pol II in red and Pol I in blue. The width of the cleft opening is indicated by a dashed line and the Cα-Cα distance across the cleft. Pol I has the widest clamp opening, followed by Pol II and Pol III.

Image from Hoffmann *et al.*, 2015

3.3. The heterodimer

The RNA polymerases I and III have two additional subunits, that form a heterodimer which is attached to the core. The Pol III subunits C37-53 are reported to be important for transcription termination (Landrieux et al. 2006). It had long been discovered that in Pol III transcription termination and release of the polymerase from the transcript were two independent steps (Campbell & Setzer 1992).

In 2006 Landrieux and colleagues created a mutant that lacked the 27 C-terminal residues of C37. Cells expressing this mutant C37 were thermosensitive, but viable. From these cells an RNA polymerase III was purified that lacked the subunits C37-C53 and C11. *In vitro* transcription assays showed that it had a faster elongation rate than the wild type, but its termination was defective. The read-through defect could be restored by adding recombinant C37-C53, however, transcription reinitiation remained defective. The authors conclude that the heterodimer C37-C53 is important for getting the polymerase to pause at the transcription site, but that it is C11 that is important for rapid reinitiation of transcription. Recently a structural study confirmed the existence of a large contact surface between C11 and C37 (Hoffmann et al. 2015). Another study observed that the C37-C53 subcomplex plays also a role in the formation of the open promoter complex (Kassavetis *et al.*, 2010).

In Pol III transcription a stretch of 5-7 thymines on the non-template strand is sufficient for transcription termination (Arimbasseri *et al.*, 2014). It has been shown that C37 possess a loop structure (residues 197-224) that reaches down in the DNA cleft then bends back into a helix (residues 230-240) (Hoffmann et al. 2015). Deleting the five residues that lead into the helix (R226, L227, T228, G229, S230) will result in a read-through mutant. This leads to the hypothesis that this helix comes into contact with the non-template strand and recognizes the thymine stretch that serves as termination signal.

3.4. The heterotrimer

Pol III possesses three additional subunits, not present in Pol I or III. These subunits, C82-C34-C31 form a detachable subcomplex positioned on the polymerase core (Wang & Roeder 1997). The same study showed that a purified Pol III lacking this subcomplex was still capable of transcription elongation and termination, but lacked the ability to initiate promoter directed transcription.

Early on, it was shown that the subunit C34 had a role in transcription initiation and open complex formation (Werner *et al.*, 1993; Brun *et al.*, 1997). Using yeast two-hybrid and pull down assays, it was observed that C34 binds to Brf1, a subunit of the transcription factor IIIB (TFIIIB) (Werner *et al.*, 1993; Khoo *et al.*, 1994). It took over a decade before these observations were confirmed *in vivo* in mammalian cells (Kenneth *et al.*, 2008).

The subunit RPC39 which is the human counterpart to yeast C34, has been found to bind double stranded DNA, whereas RPC62, the human form of C82, preferentially binds single stranded DNA (Lefèvre et al. 2011). Recently it has been shown that RPC62 possess a helical activity, which leads to believe that it plays a role in the opening of the transcription bubble (El-Ayoubi *et al.* unpublished data).

The third subunit of the complex C31 has been shown to be important for transcription initiation (Thuillier et al. 1995). The authors analyzed several C31 mutants and found that deletion of more than 16 amino acids from the C-terminus resulted in a lethal phenotype. The deletion of less than 10 amino acids led to no detectable phenotype. A mutant termed C31-236 which carried a deletion of the C-terminal 16 amino acids was thermosensitive and showed reduced doubling times at 30° C. Interestingly purified Pol III carrying the mutant subunit was still able to carry out non-specific transcription at a rate similar to that of the wild type. Transcription termination and reinitiation were not impaired either. However, specific transcription of tRNAs was reduced to 20-40% of the wild type level, depending on the tRNA. These results indicate that C31 plays an important role in promoter specific transcription initiation.

The human homolog of C31, RPC32 is unique among the Pol III subunits, as it exists in two forms: RPC32α and RPC32β (Haurie et al. 2010). Only one of the two variants will be integrated in to the polymerase leading to the formation of either Pol IIIα or Pol IIIβ. The characteristics of each of the two forms are discussed further below.

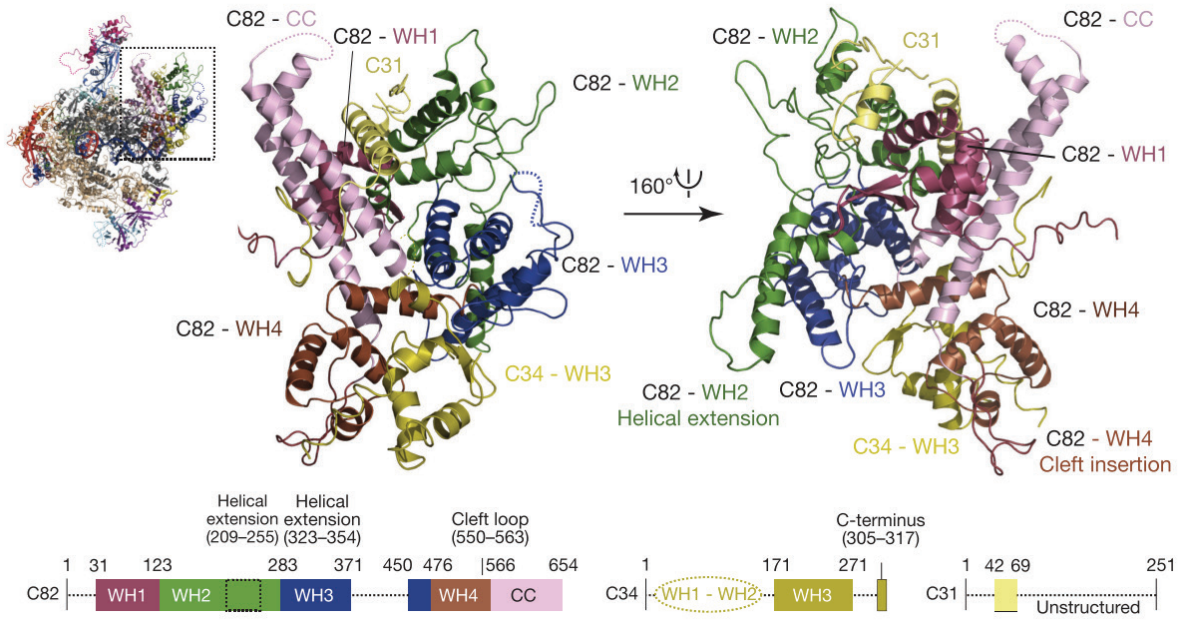

Figure 9: Architecture of the Pol III specific heterotrimer. In the upper panel the structure of the C82-C34-C31 heterotrimer is given in ribbon representation. The lower panel shows a schematic overview of structured and unstructured regions of the three proteins.

Image from Hoffmann *et al.*, 2015

On a structural level it was shown that the subcomplex C82-C34-C31 has 7 winged-helix (WH) domains, four in C82 and three in C34 (Hoffmann et al. 2015; Lefèvre et al. 2011) (figure 9). These domains are often found in protein-protein or protein-DNA interactions. Three of the WH domains, C82-WH1/WH4 and C34-WH3, are involved in tying the complex to the clamp head. The other two C82 WH domains, WH2 and WH3, are facing away from the core towards the stalk. The exact position of the two N-terminal WH domains of C34 could not be elucidated (Hoffmann et al. 2015). A previous study has described C34 as bridging the active center cleft, contacting the two major subunits C160 and C128, as well as C82 (Wu et al. 2012). The same study identified intra-subcomplex links between C82-C34 and between C82-C31. C31 is furthermore connected to C160 and RPB5 and to the stalk, making it a key element that attaches the subcomplex to the core (Ferri et al. 2000; Wu et al. 2012; Hoffmann et al. 2015).

3.5. The subunits RPC32α and RPC32β

A database search revealed the existence of a paralog of the Pol III subunit RPC32 (Haurie et al. 2010). The two proteins were termed RPC32α and RPC32β and affinity chromatography followed by western blot and mass spectrometry confirmed that both proteins can bind to other subunits of Pol III, but will not bind to each other (Haurie et al. 2010; Renaud et al. 2014). This indicates that the RNA polymerase III can alternatively be built with RPC32α or RPC32β, leading to either Pol IIIα or Pol IIIβ.

3.5.1. Evolution

The two proteins are encoded by different genes. RPC32α is encoded by POLR3G, which is located on chromosome 5, whereas RPC32β is encoded by POLR3GL (for POLR3G-like) located on chromosome 1. While the genes differ in sequence and size (40 629 bp for RPC32α versus 19 812 bp for RPC32β), the gene structure is identical. Both genes are divided into 8 exons, with the start codon in the second and the stop codon in the eighth exon. This indicates that the genes most likely resulted from a DNA rather than an RNA based duplication. On a protein level RPC32β shows 47% amino acid identity and 53% homology (Haurie et al. 2010).

A study that searched for copies of POLR3G-related genes in different genomes found that most mammals possess two copies of POLR3G, and the thirteen-lined ground squirrel (*Spermophilus tridecemlineatus)* even possess three copies (Renaud et al. 2014) (figure 10). Three copies are also found in *Danio rerio*, but other fish only have one or two copies. One single copy was identified in birds and simpler life forms such as *S. cerevisiae*, *Drosophila melanogaster* or *Ciona intestinalis*, a member of the tunicate. It was therefore concluded that the duplication took place in an ancestor of

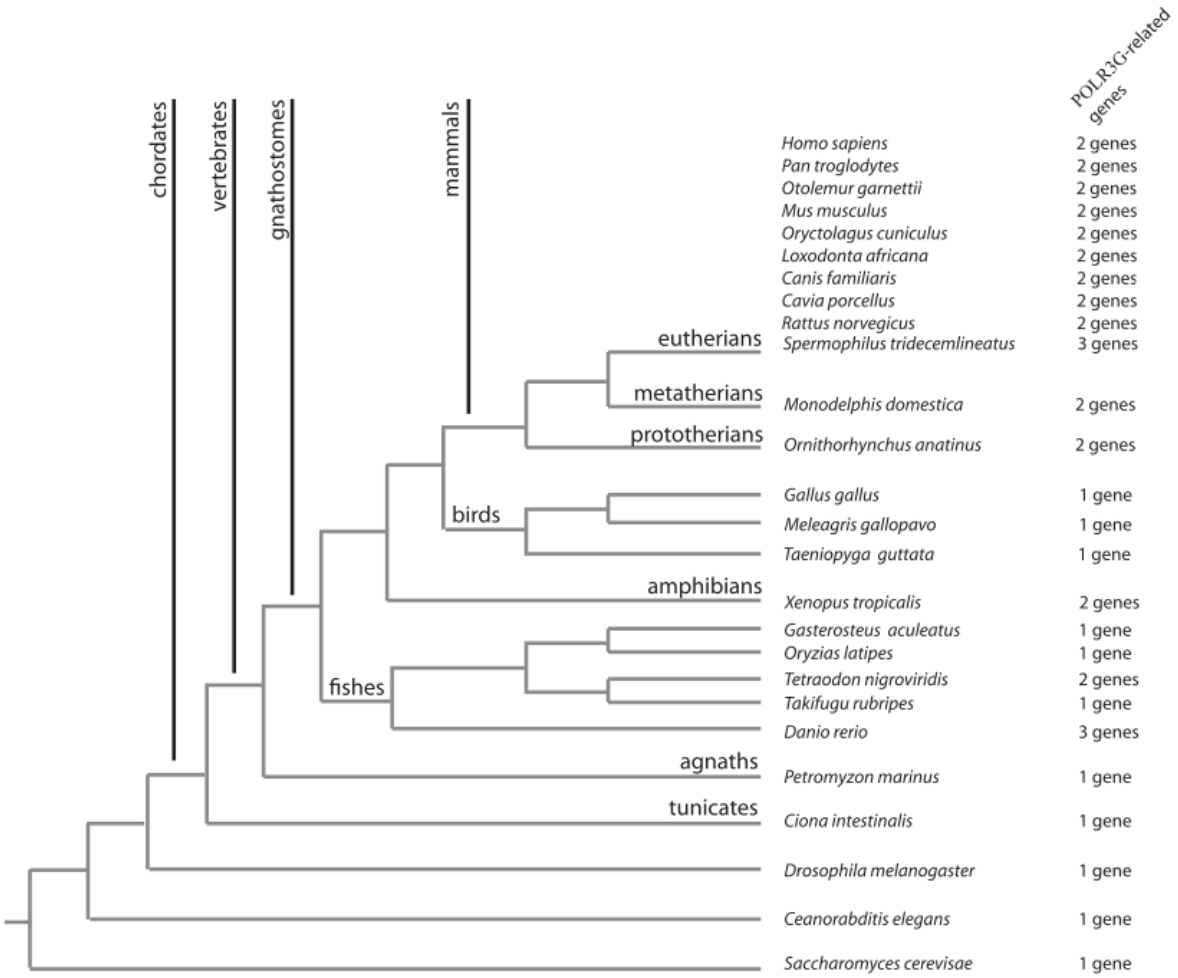

Figure 10: Number of the different POLR3G-like genes in different organisms. Most mammals and many of the gnathostomes have more than one copy of the gene. Indicating that a duplication event probably took place in an ancestor of the vertebrates, which was followed by the loss of one copy in birds.

Image from Renaud *et al.*, 2014

the vertebrates and that one copy has been lost in birds and some fish (Renaud et al. 2014). The study identified POLR3GL as being more closely related to the ancestral gene than POLR3G.

3.5.2. Structure

Recently a study reported the structure of Pol III at 3.9 Å (Hoffmann et al. 2015). The study had been done on a Pol III purified from *Saccharomyces cerevisiae*, which possess only one copy of a POLR3G-related gene. The study reports C31, the yeast homolog of RPC32, to be mostly disordered, but they identified one helical element (residues 42-69) along the surface of C34. Another study describes the C-terminus of C31 as flexible and disordered (Hoffmann et al. 2016). Recently a study attempted to analyze the structure of the human RPC32 faced the same difficulties (Boissier et al. 2015). The authors describe RPC32 as belonging to the class of proteins that are natively disordered. These

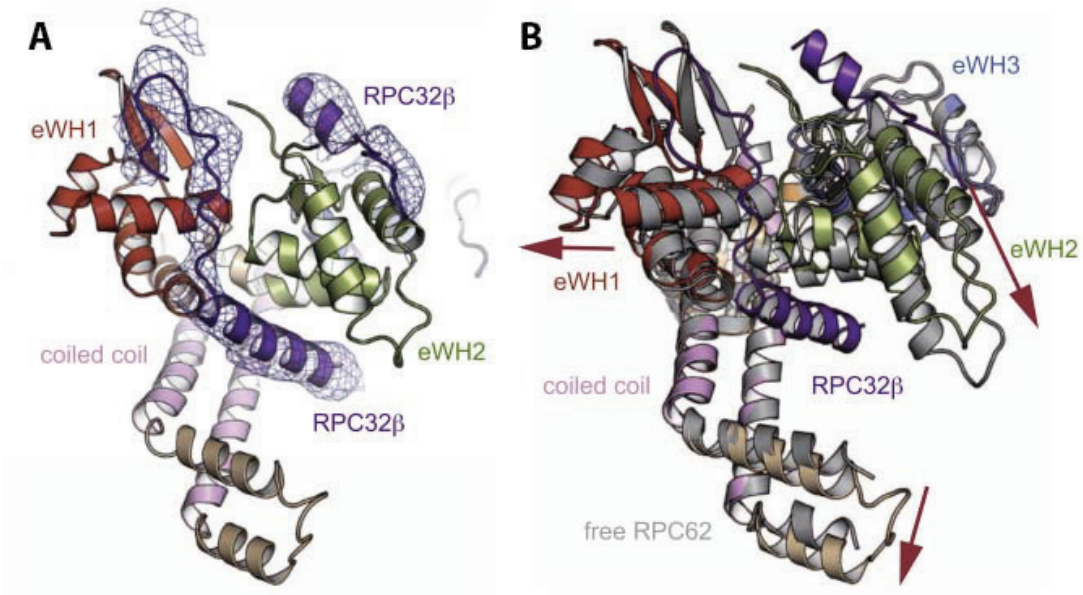

Figure 11: Structure of the RPC32β- RPC62 complex. (A) RPC32β interacts with eWH1 and eWH2 of RPC62. (B) The superposition of the structure of RPC62 alone and the RPC32β-RPC62 complex reveals rearrangements of RPC62 when it is bound to RPC32β. Upon binding the two winged helices eWH1 and eWH2 are displaced in the direction of the red arrows by 4Å and 6Å respectively. The position of eWH2 and eWH4 remains mostly unchanged. Furthermore, binding to RPC32β also affects the helices of the coiled coil of RPC62, which are shifted away from the core by 2 – 4Å (red arrow).

Image adapted from Boissier *et al.*, 2015

proteins obtain their secondary structure either on binding to their natural partner or they serve as scaffolds to stabilize inter-subunit contacts. To circumvent these difficulties the authors tried to purify RPC62-RPC32α/β dimers. They obtained crystals only for the RPC62-RPC32β dimer, which led to a low-resolution structure (7.38-7.0 Å). The structure shows that RPC32β stabilizes and reorients RPC62 (figure 11). Furthermore, the authors suggest that RPC32β is exposed at the surface of Pol III and that any functional difference between RPC32α and RPC32β may reside in the N- and C-terminal extensions (Boissier et al. 2015).

3.5.3. Expression

The two genes coding for the two RPC32 paralogs show distinctive expression patterns. POLR3G, which codes for RPC32α, is among the most highly upregulated genes in human embryonic stem cells (hESC) compared to differentiated cells (Enver et al. 2005). Upon differentiation the expression level of POLR3G drops rapidly, while that of POLR3GL remains stable (Haurie et al. 2010). Dot blot analysis revealed that POLR3GL, which codes for RPC32β, is expressed in a great variety of tissues, with the lowest expression levels in tumors. POLR3G on the other hand is barely detectable in differentiated tissues, but showed high expression levels in tumors (Haurie et al. 2010). A recent study analyzed the expression levels of RPC32α and RPC32β in a tumor model system, using fibroblasts that had been transformed with defined genetic elements (Durrieu-Gaillard *et al.*, under revision). It was shown

that as the cells became more transformed, the expression level of RPC32α increased, while the level of RPC32β decreased.

Two studies have analyzed the effects of a knock-down of POLR3G in tumor and stem cells. In tumor cells the siRNA-mediated suppression of RPC32α did not lead to a change of growth rate in Petri dishes, however, it reduced anchorage independent growth, as shown by soft agar assays (Haurie et al. 2010). In stem cells the reduction of POLR3G caused the cells to differentiate (Wong et al. 2011). The knock-down of POLR3GL has only been characterized in tumor cells, where it led to substantially reduced cell growth and increased cell death (Haurie et al. 2010). No study has so far analyzed the effects of a complete knock-out of either RPC32α or RPC32β.

Overexpression of RPC32α in fibroblasts let to increased growth rates and induction of anchorage independent growth (Haurie et al. 2010). Overexpression of RPC32β in the same type of cells led to slight inhibition of soft agar colony formation. In stem cells the ectopic expression of RPC32α has no visible effect on the cells, but renders hESC more resistant to spontaneous differentiation (Wong et al. 2011). Interestingly the overexpression of RPC32α in stem cells did not lead to a change of the level of endogenous RPC32α, indicating that no auto-regulatory mechanism exists.

3.5.4. Function

The fact that two paralog subunits exist poses the question of if and how they might differ functionally. To test the capacity of Pol IIIα and Pol IIIβ to transcribe different Pol III promoters, Haurie *et al.* (2010) purified Flag-tagged RPC32α and Flag-tagged RPC32β from HeLa cells. Together with other factors from the Pol III transcription machinery (recombinant Bdp1, rBrf1, rTBP, Flag-TFIIIC, rPCF4) the two Pol IIIs were used for *in vitro* transcription assays. It was shown that both polymerases were able to transcribe the genes of 5S, tRNA, VA1 and 7SK. Thus demonstrating that both Pol IIIα and Pol IIIβ are capable of transcription from all known Pol III promoter types.

Following overexpression of RPC32α, the authors analyzed a handful of Pol III transcripts *via* qPCR (Haurie et al. 2010). While there had been a large increase in U6 snRNA, 5S rRNA and 7SK RNA, the increase was only moderate for $tRNA^{met\ i}$ and BC200. No increase was observed for $tRNA^{Glu}$ or vault 1 RNA. None of these transcripts showed a change in expression levels for overexpressed RPC32β (Haurie et al. 2010). In stem cells the overexpression of RPC32α did not lead to a change in Pol III transcript expression. The knock-down of RPC32α however resulted in an increase of $tRNA^{Leu}$ and 5S rRNA, while it led to a decrease in 7SL RNA (Wong et al. 2011). Furthermore, the downregulation of RPC32α also resulted in a reduction of mRNA for OCT4, SOX2 and NANOG in hESC (Wong et al. 2011). This is conform with the cells losing pluripotency upon knock-down of RPC32α.

To identify potential differences in transcription between Pol IIIα and Pol IIIβ, Renaud *et al* (2014) performed ChIP-seq analysis on IMR 90 cells. Both RPC32 forms occupied genes with type 1, 2 or 3 promoters. No clear preference of either RPC32α or RPC32β for a certain type of transcript could be identified. In a subsequent analysis Renaud *et al* (2014) compared ChIP-seq data from mouse liver and hepatocarcinoma cells. The authors noticed a general increase of Pol III occupancy in the cancer cells compared to normal cells, which is consistent with the idea that Pol III has increased activity in cancer cells (White 2004; Johnson et al. 2008). The increase in Pol III occupancy resulted from an increase in RPC32α, but all loci were occupied by both RPC3α and RPC32β. No locus was significantly associated with just one of the two paralogs.

3.5.5. Regulation

All studies on RPC32α and RPC32β report that the two paralogs are differentially expressed, so they must be differentially regulated. While RPC32α is strongly expressed in some tumors and stem cells, RPC32β is present in fibroblast and a variety of differentiated tissues (Haurie et al. 2010; Wong et al. 2011; Renaud et al. 2014). But the expression levels do not only vary with tissue types, but also with culture conditions such as starvation or confluence. Even if the reports here are not univocal. Haurie *et al* (2010) report a 2.6-fold increase of POLR3G upon starvation in IMR90 fibroblast cells. While Renaud *et al* (2014) observed a decrease in POLR3G expression in serum starved IMR90Tert cells. The difference might be explained by the nature of the cells, as IMR90 cells are fibroblasts, while IMR90Tert cells have been immortalized by the ectopic expression of the human telomerase reverse transcriptase (TERT).

Several regulatory pathways have been linked to POLR3G. Wong *et al* (2011) identified OCT4 and NANOG as upstream regulators of POLR3G. Furthermore, they could show that inhibition of the extracellular signal-regulated kinase 1/2 (Erk1/2) lead to a decrease in POLR3G expression. Inhibition of the mammalian target of rapamycin (mTOR) or the phosphatidylinositol 3-kinases (PI3K) did not affect POLR3G levels. Therefore POLR3G is likely part of the Erk1/2 pathway, but not the mTOR or PI3K pathway (Wong et al. 2011). Recently two miRNAs were identified that regulate the expression of POLR3G: miR-27 and miR-1305 (Fuchs et al. 2014; Jin et al. 2016).

Renaud *et al* (2014) identified MYC binding sites close to the TSS of POLR3G, but not POLR3GL. Moreover, they found MYC binding sites close to all genes of the other 16 subunits of Pol III, making POLR3GLs lack of a binding site a real exception. MYC is a transcription factor, which is constitutively active in many cancers (Eilers & Eisenman 2008). Indeed ectopic expression MYC led to accumulation of MYC and Pol II at the TSS of POLR3G, but not of POLR3GL (Lin et al. 2012; Renaud et al. 2014). If

POLR3G and the rest of Pol III are activated by MYC, but not POLR3GL, this would explain why Pol III in cancer cells is mostly in form of Pol IIIα (Renaud et al. 2014).

4. Pol III and cancer

More than 40 years ago scientists discovered that Pol I and III were hyperactive in myeloma cells, while Pol II transcription levels remained the same (Schwartz et al. 1974). This was the first indication that Pol III activity might be linked to tumorigenesis. Since then numerous articles have described deregulated Pol III activity in cancer cells (Tang et al. 2005; Winter et al. 2000; Johnson et al. 2008; Lee 2015; Booy et al. 2017). It was shown that Pol III transcripts were deregulated in breast, cervix, esophagus, lung, ovary, parotid, and tongue tumors, but not in the corresponding healthy tissue (Chen et al. 1997). But whether this deregulation was a cause or a consequence of tumorigenesis remained unclear.

An answer to this question came in 2008 (Johnson, *et al.*, 2008). A study analyzed Rat1a fibroblast during oncogenic transformation. The authors observed that transformed cells had elevated Pol III transcription levels. Using shRNA they reduced the levels of Brf1, one of the subunits of TFIIIB, thus limiting Pol III recruitment to its promoters. As expected the transcription levels of Pol III came down in the transformed cells. While the cells showed unchanged proliferation rates, they had lost the ability for anchorage independent growth, one of the hallmarks of tumorigenesis. This indicates that increased Pol III transcription is not a consequence, but a necessity for tumorigenic growth.

Given its importance for transformation and oncogenesis, it does not surprise that Pol III activity is closely regulated by several tumor suppressors and oncogenes. One of the most well characterized tumor suppressors, p53, is a repressor of Pol III activity. The first connection between p53 and Pol III was made, when it was discovered that p53 could reduce the level of Alu element transcription both *in vitro* and *in cellulo* using wild type and mutant p53 cells (Chesnokov et al. 1996). In fibroblasts that were derived from p53 knock-out mice, it was discovered that p53 is a general inhibitor of Pol III activity (Cairns & White 1998). The inhibition through p53 is not dependent on DNA binding or cell cycle regulation. Rather p53 binds to the TFIIIB subunit TBP and prevents its interaction with TFIIIC and recruitment of Pol III to its promoters (Crighton et al. 2003).

Another famed tumor suppressor, the retinoblastoma protein (Rb), also controls the activity of Pol III. A first indicator for a regulation *via* Rb was the observation that Pol III activity is diminished during G1 phase of the cell cycle and increases shortly before the transition to the S-phase (White et al. 1995). This is inversely proportional to the activity of Rb. Shortly later it was found that $Rb^{-/-}$ mice had high levels of Pol III activity and that overexpression of Rb would lower Pol III transcription levels (White et al. 1996). Like p53, Rb also interacts with TFIIIB, notably with its subunits BRF1 and TBP

(Larminie et al. 1999). As BRF1 is only required for the transcription from type 1 and 2 promoters, the question arose whether additional targets existed in the case of type 3 promoters. Indeed in 2004 a study identified interactions between Rb and type 3 transcription factors, among them SNAPc (Hirsch et al. 2004).

Besides p53 and Rb, a number of other proteins act as suppressors to Pol III activity. The pocket proteins p107 and p130 are closely related to Rb and can too repress Pol III transcription (Sutcliffe et al. 1999). Known tumor suppressor PTEN negatively regulates Pol III by disrupting the association of TBP and BRF1, thereby limiting the number of correctly assembled TFIIIB units (Woiwode et al. 2008). It furthermore counteracts the PI3K pathway, which normally favors Pol III transcription (Woiwode et al. 2008).

While tumor suppressors inhibit the activity of Pol III, oncogenes favor it (figure 12). As mentioned above, the oncogene MYC can bind directly to the promoters of all Pol III subunits except POLR3GL (Renaud et al. 2014), thus promoting directly Pol III subunit transcription. But ChIP-seq revealed that MYC is also present at the TSS of Pol III transcripts such as tRNAs (Gomez-Roman et al. 2003). MYC had not been suspected at these sites, as they do not possess the MYC binding sequence CANNTG. It was shown that MYC is present at POL III transcript sites, due to protein-protein interactions with TFIIIB (Gomez-Roman et al. 2003).

During mitosis, MYC promotes an increase in the expression of a vast number of genes (Naldini et al.

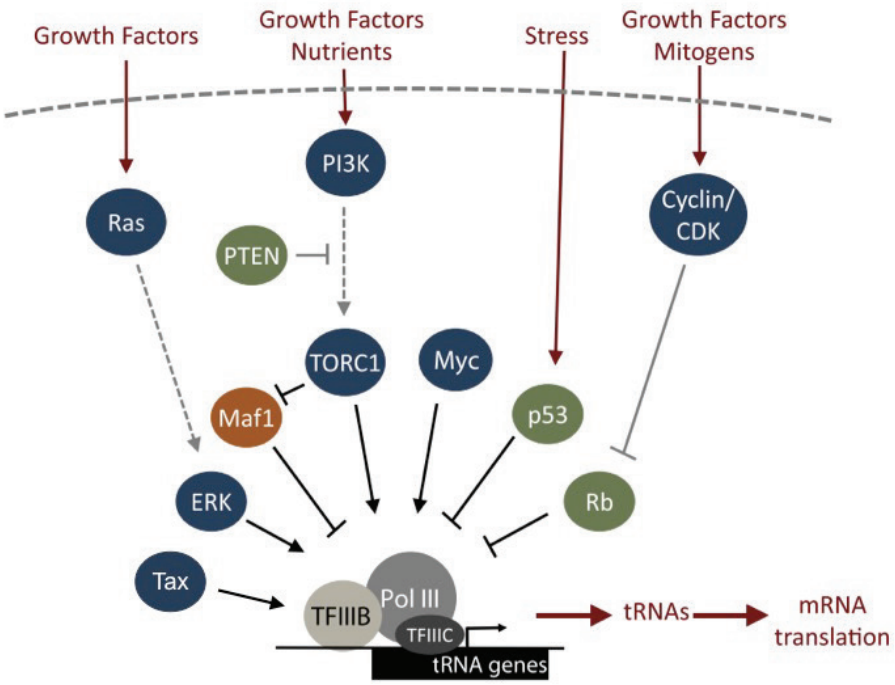

Figure 12: The regulatory network of Pol III. The activity of the RNA polymerase III is tightly controlled. While tumor suppressors (green) inhibit Pol III, oncogenes (blue) stimulate it. Regulation of Pol III Is important as its transcripts, such as tRNAs are involved in important cellular processes including mRNA translation.

Image adapted from Grewal, 2014

1996). However, MYC-driven gene expression in tumor cells is limited to a specific set of genes (Horiuchi et al. 2012; Yustein et al. 2010; Valentijn et al. 2012). To elucidate the mechanisms behind this specificity, a study analyzed the effect of MYC expression in doxycycline inducible U2OS bone sarcoma cells, which normally have low MYC levels (Walz et al. 2014). The authors found that MYC occupancy was enriched at promoters of Pol II and Pol III. Furthermore, they noted that MYC can both up- and downregulate gene expression. The same was true in HeLa cells, in which MYC had been downregulated. The authors find that genes that have a canonical CACGTG MYC binding site, tend to be downregulated. Genes with a non-canonical CANNTG binding site, were more likely upregulated (Walz et al. 2014). The promoters of the Pol III subunits, except POLR3GL, have a CANNTG binding site (Renaud et al. 2014).

Besides direct interaction with the Pol III transcription apparatus, MYC also activates Pol III transcription indirectly. Once bound to TFIIIB, MYC will help recruit the histone acetyltransferase CGN5 and its cofactor TRRAP, which further enhance transcription (Kenneth et al. 2007). A lesser-known oncogene is the human T-cell leukemia virus type 1 Tax protein. It was shown that HTLV-1 Tax stimulates Pol III transcription through interaction with TFIIIB (Gottesfeld, *et al.*, 1996).

Pol III is also implicated in different cell signaling pathways that often get deregulated in cancer. The target of rapamycin (TOR) is a serine/threonine kinase in the PI3K pathway. In most eukaryotes TOR can form two functional distinct complexes TORC1 and TORC2, which are both involved in regulating cell growth (Betz & Hall 2013). Almost 20 years ago it was shown that Pol III is enhanced by TOR (Zaragoza et al. 1998). Further studies showed that TOR has both a direct and indirect positive effect on Pol III transcription. Direct interactions of TOR and BRF are necessary for Pol III activity in drosophila and inactivation of TORC1 through PTEN leads to reduced numbers of TBP-BRF complexes (Zhang et al. 2000; Woiwode et al. 2008). Most studies report however on the indirect activation of Pol III, with TOR as an inhibitor of the Pol III repressor MAF1.

Maf1 was first identified as Pol III suppressor in yeast, where it inhibits the assembly of TFIIIB onto the DNA (Upadhya *et al.*, 2002). As Maf1 is highly conserved from yeast to humans, it came as no surprise when its repressor function was confirmed for human cells (Reina et al. 2006). MAF1 is phosphorylated in human cells, but can be dephosphorylated upon stress. The dephosphorylated form will inactivate Pol III, until MAF1 gets phosphorylated again and releases Pol III (Reina et al. 2006). Among the kinases that have been proven to phosphorylate MAF1 is mTOR (Kantidakis et al. 2010).

The Ras family of proteins is a class GTPases. These proteins function like a switch and can transmit signals to a cell. In many cancers Ras is mutated and constitutively active (Pylayeva-Gupta, *et al.*, 2011). Through different signaling pathways Ras is linked to a large number of cellular processes, such as cytoskeletal organization, survival, proliferation, vesical trafficking and calcium signaling

(Schubbert, *et al.*, 2007). *Via* the MEK/ERK cascade Ras is also linked to proliferation. ERK1/2 has been shown to favor Pol III transcription by phosphorylating BRF1 and thereby enhancing its interaction with TFIIIC (Felton-Edkins & White 2002). ERK activation is also required for the increased expression of TBP and BRF1 (Zhong *et al.*, 2004; Goodfellow *et al.*, 2006). For stem cells it has been shown that the ERK1/2 pathway positively regulates the expression of POLR3G, gene that codes for RPC32α (Wong et al. 2011).

5. Breast cancer

According to the GLOBOCAN project, initiated by the World Health Organization (WHO), breast cancer is the second most common cancer worldwide, with lung cancer being number one (GLOBOCAN12: Ferlay *et al.*, [Internet]). Among women, breast cancer is by far the most prevalent cancer in both developed and developing countries. A total of 1.68 million new cases were diagnosed worldwide in 2012, which represents 12% of all new cancer cases and 25% of cancers in women. However, the incidence rate varies greatly across countries, from 19.3 cases per 100 000 women in Eastern Africa to 89.7 cases in Western Europe. The large discrepancy is partly due to a lack of cancer detection in developing nations, nonetheless there seems to be a bias towards western countries.

But even if women in developed countries are more likely to get breast cancer, they are also more likely to survive it (figure 13). A woman being diagnosed with breast cancer in the USA, Japan or France has a twice better chance of survival than a woman living in Algeria (Coleman et al. 2008). But the high survival rates in western countries depend largely on early discovery of the disease. While the 5-year survival rate is about 80-90% for early stage cancers, it drops to 24% for cancers diagnosed at a more advance stage (GLOBOCAN12: Ferlay *et al.*, [Internet])

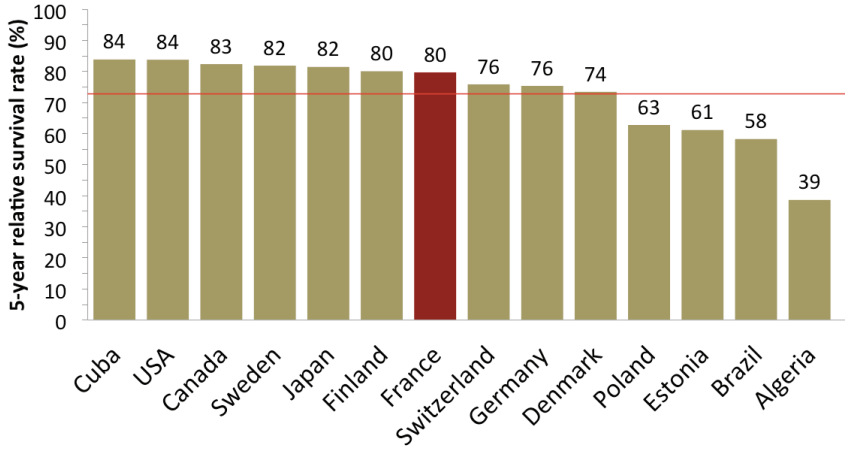

Figure 13: 5-year survival rate of breast cancer patients by country. The chances of surviving breast cancer are very unequal around the world. The best chances of survival have patients in Cuba, the worst women living in Algeria. The red line indicates the mean survival rate of women in Europe.

Data from Coleman *et al.*, 2008

5.1. Risk factors

The factors that lead to a higher risk can be divided into reproductive, behavioral and genetic factors. Among the reproductive factors are early menarche, late menopause and late age at first child birth. All of these factors lead to a prolonged exposure to estrogen and progesterone, which increases the risk of breast cancer (GLOBOCAN12: Ferlay *et al.*, [Internet]). For the same reason the use of hormonal products during hormone therapy or in form of oral contraceptives can raise the risk of breast cancer (WCRF). These factors are often associated with a western lifestyle, which partly explains why the incidence rate in developed countries is so much higher.

Other factors that will favor the onset of breast cancer are alcohol abuse, obesity and physical inactivity (GLOBOCAN12: Ferlay *et al.*, [Internet]). A study that evaluated these behavioral factors concluded that they are responsible for 21% of all breast cancers worldwide (Danaei et al. 2005). In high-income countries 27% of breast cancers are due to such behavioral factors, with the most important risk factor being obesity. In low- and middle-income countries only 18% of all breast cancer were attributed to these factors. Here the prevailing risk factor was physical inactivity (Danaei et al. 2005).

About 5-10% of all breast cancers are hereditary, that is they follow an autosomal dominant Mendelian inheritance pattern (Siegel, *et al.*, 2013). Another 15-20% of all cases are familial breast cancer, meaning the patient has two or more first or second degree relatives who have the disease (Siegel, *et al.*, 2013). The more relatives are affect and the younger they were at the onset of cancer, the more it is likely that a person will develop breast cancer (Lalloo & Evans 2012).

Among the hereditary cancers about 30% are due to a mutation in the genes breast cancer 1 (BRCA1) or breast cancer 2 (BRCA2) (Miki et al. 1994; Hall et al. 1990; Wooster et al. 1995; Ford et al. 1998). The two genes code for proteins implicated in DNA repair. Other genes that are frequently found mutated in hereditary breast cancers include p53, the serine/threonine kinase 11 (STK11), the phosphatase and tensin homolog (PTEN), Cadherin 1 (CDH1) and genes involved in DNA mismatch repair (MMR genes). Besides these so called high penetrance genes, a number of other genes are found to be mutated in familial breast cancer, including the checkpoint kinase 2 (CHEK2), Ataxia-telangiectasia (ATM), the partner and localizer of BRCA2 (PALB2) and the BRCA1 Interacting Protein C-Terminal Helicase 1 (BRIP1). With the expansion of next generation sequencing, more and more gene mutations are identified, but more research needs to be done to identify the risk associated with each mutation.

5.2. Diagnosis

The female breast is composed of fatty disuse and milk glandes. The latter are formed of alveoli, which join together to form lobules, which in turn group together in lobes (figure 14 A). The milk is produced in the alveoli and flows through milk ducts towards the nipple. The ducts have an outer layer of myoepithelial cells, surrounded by a basal membrane (figure 14 B and C). The mammary milk gland evolves through distinct stages, namely puberty, pregnancy, lactation and involution. All these developmental stages are regulated by hormones, which will induce the mammary stem cells (MSC) to differentiate into either epithelial or myoepithelial cells.

The vast majority of breast cancers develops from epithelial cells in the ducts or the lobules. These cancers are very heterogeneous and different classifications are used to diagnose them. A correct diagnose is especially important as it will impact the kind of treatment that is to be followed. For instance, cancers can be rated according to tumor stage, tumor grade, hormone receptor status and certain molecular markers.

By biopsy and subsequent histological analysis, the general type of breast cancer is determined. It can either be *in situ*, meaning the cancer cells have not yet spread to other tissue (figure 15). Depending on the site of the cancer, one discriminates between *ductal carcinoma in situ* (DCIS) or *lobular carcinoma in situ* (LCIS). If the cancer cells have spread to the surrounding tissue, the cancer is classified as invasive. Again depending on the site, cancers are classified as invasive ductal

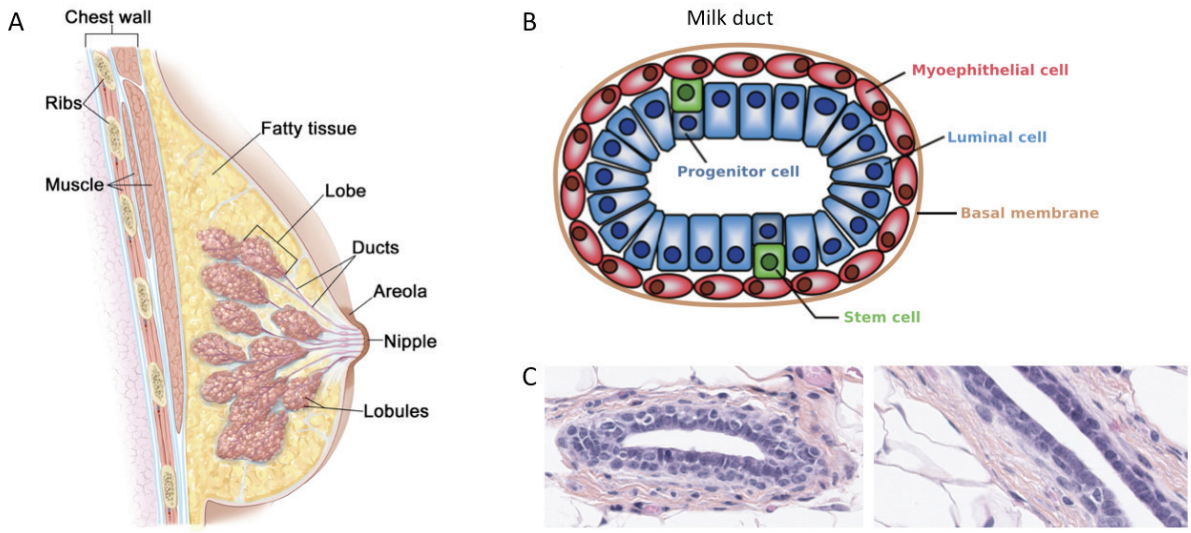

Figure 14: Anatomy of the female breast. (A) The female breast is composed of fatty tissue and milk producing lobes. Each lobe is formed of several smaller lobules, which secret milk. The milk then flows through the ducts toward the nipple. (B) Schematic cross section of a milk duct. The milk duct is formed of luminal milk secreting cells, which are surrounded by myoephitelial cells and a basal membrane. Inside the milk duct mammary stem cells, which can differentiate into either luminal epithelial or myoephithelial cells. (C) Histological view of a milk duct from the front (left) and side (right).

Images adapted from www.cancer.org and Tiede and Kang, 2011

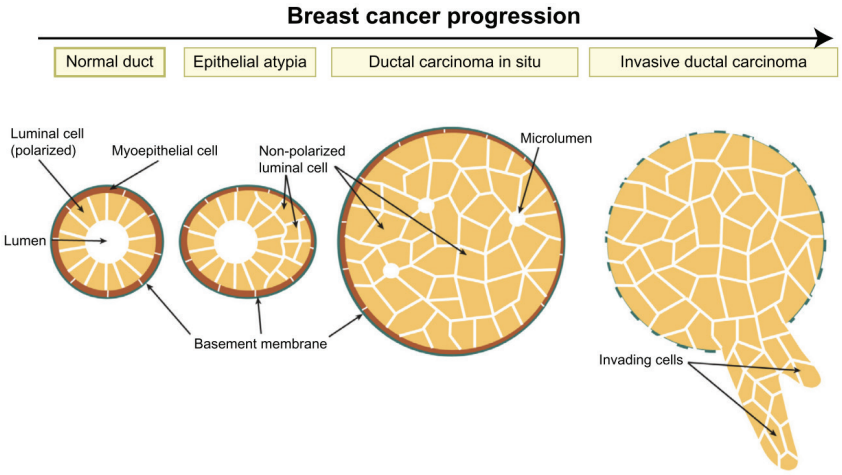

Figure 15: Evolution from a normal milk duct to an invasive ductal carcinoma. The normal milk duct is a well defined structure with an inner lumen for milk flow. During carcinogenesis the inner lumen is filled up with non-polarized luminal cells. As long as the outer membrane is intact and the tumor is contained in the milk duct, the cancer is classified as *ductal carcinoma in situ*. Once the basal membrane is lost and the cancer starts invading other tissues, the tumor becomes an *invasive ductal carcinoma.*

Image from Chatterjee and McCafferey, 2014

carcinoma (IDC) or invasive lobular carcinoma (ILC). About 80% of all cancers are DCIS and five subtypes (tubular, medullary, mucinous, papillary and cribriform) are used to further classify this vast and heterogeneous group.

Furthermore, tumors are staged, depending on the size of the cancer, whether or not it is invasive and if it has spread to the lymph nodes. Tumor stage can range from 0 to IV, with 0 indicating that the tumor has not invaded surrounding tissue and IV that the cancer has invaded other body parts like the lungs, the liver or the brain. All *in situ* carcinoma are by definition stage 0. The invasive carcinoma can range from I to IV. The higher the stage, the larger the main tumor and the more it has spread to surrounding tissue. Stage IV tumors have spread to distant parts of the body. A widely used system is the TNM staging system, which looks at the primary tumor (T), the lymph nodes (N) and metastasis (M).

To determine how likely a tumor is to grow and spread, it is graded according to the cells phenotype. The grading system differs for different cancers. In breast cancer three parameters are taken into account: tumor structure, size and form of the cell nucleus and the percentage of dividing cells present. The combination of these parameters makes the tumor grade, which can vary from 1 to 3. Grade 1 tumors have well differentiated cells, small uniform nuclei and low levels of diving cells. Grade 3 tumors on the other hand have undifferentiated cells, large nuclei of variable form and a high number of dividing cells. The higher the assigned grade is, the poorer the predicted prognosis.

Another important step in diagnosing cancer is the identification of its hormone receptor status. Among the receptors tested are the receptors for estrogen (ER), progesterone (PR) and the human epidermal growth factor receptor 2 (HER2). Cells that test positive for any one of these receptors can

be targeted with hormonal therapy or monoclonal antibodies. Cells that lack all three receptors are called triple-negative.

5.3. Molecular subtypes of breast cancer

Until the late 1990s, histological assessments were the only option to classify breast cancers. But with the advent of new biomolecular techniques, such as micro arrays, new characterizations have emerged. The first genetic studies confirmed the heterogenic complexity of breast cancers and established several main subgroups (Perou et al. 2000; Sorlie et al. 2003). The new molecular classification together with the established histological analysis, established five major breast cancer subtypes: luminal (A and B), $HER2^+$, triple-negative breast cancer (basal and non-basal).

5.3.1. Luminal cancers

All cancers that test positive for the estrogen and/or the progesterone receptor [ER^+|PR^+] are called luminal cancers. The name is derived from the fact that the cells show expression profiles reminiscent of epithelial cells found in the inner (luminal) part of the mammary ducts (Perou et al. 2000). Luminal tumors make up for about 60-70% of all breast cancer and they can be divided into two subgroups: luminal A and luminal B (reviewed in Dai *et al.*, 2015). While both subgroups are ER^+ they differ mainly in their levels of the proliferation marker Ki67 and the levels of HER2 (table 4).
Luminal A tumors are described as [ER^+|PR^+] $HER2^-$ $Ki67^-$ (Dai et al. 2015). For luminal B two subgroups have been identified: luminal B ($HER2^+$) with a [ER^+|PR^+] $Ki67^+$ $HER2^+$ signature and

Table 4: Molecular subtypes of breast cancer

Subtype	Alias	Biomarker status	Outcome	Additional features
Luminal	Luminal A	[ER+|PR+] HER2- Ki67-	Good	Luminal cytokeratin+, FOXA1+, ADHB1 high, cell-cell adhesion genes high
	Luminal B	[ER+|PR+] HER2- Ki67+ [ER+|PR+] HER2+ Ki67+	Intermediate |Poor	Luminal cytokeratin+, TP53-, ADHB1 low, cell-cell adhesion genes low
HER2 +	HER2	ER-PR-HER2+	Poor	TP53-, GRB7 high, cell-cell-adhesion genes high
Triple-negative	Basal-like	ER-PR-HER2-, basal marker+	Poor	BRCA1-, TP53-, CDKN2A high, RB low, cell-cell adhesion genes high
	Claudin-low	ER-PR-HER2-, EMT marker+, stem cell marker +, claudin-	Poor	Claudins low, CDH1 low, cell-cell adhesion genes low
	MBC	ER-PR-HER2-, EMT marker+, stem cell marker +	Poor	PIK3CA-, AKT- or KRAS-, cell-cell adhesion genes low
Molecular apocrine		ER-PR-AR+	Poor	KI67+

luminal B ($HER2^-$) or luminal B like with a $[ER^+|PR^+]$ $Ki67^+$ $HER2^+$ profile (Inic et al. 2014). But further molecular profiling has shown that these groups are still very heterogeneous and new subgroups might emerge (Yanagawa et al. 2012).

In general luminal cancers have a good prognosis with luminal A subtypes having the most favorable prospects (Sorlie et al. 2003). Treatment options for luminal cancers include hormone therapy. One of the most common medications given to luminal breast cancer patients is Tamoxifen. It blocks estrogen receptors, thereby interrupting growth signaling in cancer cells. Women before menopause can also be treated with luteinizing hormone (LH) inhibitors, which will block the production of estrogens from the ovaries. Women after menopause can benefit from aromatase inhibitors, which block the transformation from androgen into estrogen.

While luminal A tumors can be successfully treated with hormone therapy, luminal B tumor patients often receive a combination of chemotherapy and hormone blockers (Brenton et al. 2005). In some cases targeted treatments against cyclin-dependent kinases (CDK) 4/6, the vascular endothelial growth factor (VEGF) or mTOR pathway can be successful (Brenton et al. 2005; Abbas et al. 2007; Morikawa & Henry 2015).

5.3.2. HER2 positive cancers

As the name indicates $HER2^+$ tumors are characterized by an overexpression of the human epidermal growth factor 2 (HER2) additionally they have an ER^-, PR^- profile (Perou et al. 2000). Initially receptor status was analyzed by immunohistochemistry (IHC), later microarrays were developed to test for different subtypes. In the case of $HER2^+$ tumors these two methods do not always come to the same conclusions. Only 70% of intrinsic $HER2^+$ tumors, those identified by microarrays, actually show an overexpression of HER2 on a protein level (Yersal & Barutca 2014).

All $HER2^+$ tumors tend to be aggressive and have a poor prognosis (Sørlie et al. 2001; Yersal & Barutca 2014). While $HER2^+$ cancers are sensitive to chemotherapy, they show a high risk of early relapse (Brenton et al. 2005). Tremendous progress came in form of the anti-HER2 monoclonal antibody, Trastuzumab. HER2 is a tyrosine kinase that is part of a signaling pathway that activates growth. In cells that overexpress HER2 the receptor can dimerize in absence of a ligand and thus induce uncontrolled growth (Yersal & Barutca 2014).

Trastuzumab will bind and inactivate HER2 homodimers, but cannot effectively inactivate HER2 heterodimers (Ghosh et al. 2011). In 2012 Genetech announced that the Food and Drug Administration of the United States had allowed the use of Pertuzumab, a monoclonal antibody that targets HER2 heterodimers. Pertuzumab thus made possible the treatment of cancers that are insensitive to Trastuzumab.

5.3.3. Triple-negative breast cancers

Tumors that test negative for estrogen and progesterone receptors and also do not overexpress HER2 are called triple-negative tumors (table 4). They are characterized by genetic complexity, high tumor grades and poor survival rates. Between 12-17% of all cancers are triple-negative breast cancers (Foulkes *et al.*, 2010). As they do not present any hormone receptors they can neither be treated by hormone therapy nor by monoclonal antibodies. The only treatment options are chemo- and radiation therapy.

Triple-negative breast cancers are more aggressive than other breast cancer subtypes. Compared to other cancers they occur more often in young patients. The relative young age of patients and the rapid growth of triple-negative cancer makes detection using mammography difficult (Foulkes *et al.*, 2010). Furthermore, patients have a higher risk of relapse and death during the first 5 years, afterwards the risk is similar to those of other cancer types (Dent et al. 2007) (figure 16). Triple-negative breast cancers are a very heterogeneous group. Through gene expression profiling *via* micro-arrays three subgroups have been identified: basal, claudin-low and metaplastic breast cancer (MBC).

5.3.3.1. Basal cancers

Often times the term triple-negative breast cancer is used synonymously with basal-like breast cancer. The name basal cancers stems from the fact that their molecular gene expression profile

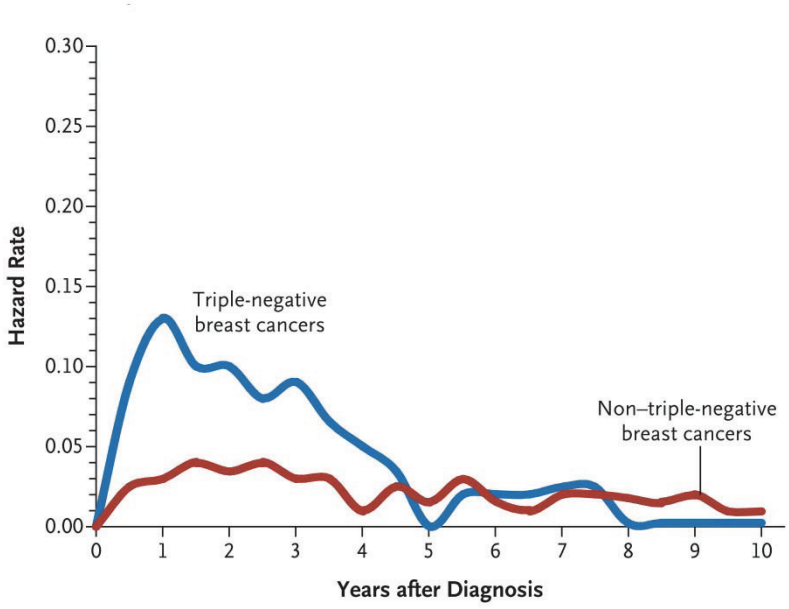

Figure 16: Hazard rate of reoccurrence. Patients that suffered from triple-negative breast cancer have a higher rate of reoccurrence during the first 5 years after treatment. Later the risk is about the same as that of non-triple-negative breast cancer patients.

Image from Foulkes, Smith and Reis-Filho, 2010

resembles those of the basal or myoepithelial cells of the breast. But even if the majority of triple-negative breast cancers are basal-like, there is about 20-30% discrepancy between the two groups (Foulkes *et al.*, 2010). The term triple-negative cancer refers to tumors that have an ER^-, PR^-, $HER2^-$ profile using immunohistochemical markers. Basal-like cancers on the other hand were identified *via* microarray analysis (Perou et al. 2000).

Among the genes typically overexpressed in basal-like breast cancers are cytokeratines (CK) 5, 14 and 17, as well as P-cadherin, vimentin, αB crystalline, fascine and caveolins 1 and 2 (Reis-Filho & Tutt 2007). Furthermore, basal-like tumors make-up for almost three quarters of all BRCA1 related cancers (Badve et al. 2010). So far there is no specific international consensus of how to define basal-like breast cancers (Yersal & Barutca 2014). Given this ambiguity and the fact that some basal-like breast cancers are ER^+ or $HER2^+$ and others not, the term basal-like has no diagnostic value. In the clinical use only the definition triple-negative is of meaning (Foulkes *et al.*, 2010).

5.3.3.2. Claudin-low and MBC cancers

The claudin-low and MBC subtypes are closely related. Both share similar tumor characteristics, genetic expression levels and clinical outcomes (Creighton et al. 2009; Hennessy et al. 2009; Prat & Perou 2011). Notably they both show high levels of stem-cell markers, such as $CD44^+/CD24^-$, which lead to the hypothesis that they originated from cells that are precursor to luminal and basal cells (Hennessy et al. 2009).

As the name indicates claudin-low tumors are characterized by low expression levels of tight junction proteins claudin 3, 4 and 7. One of the differences between claudin-low and MBC tumors are the level of mutations in the genes PI3KCA, AKT or KRAS. While these occur frequently in MBC tumors, they are rarely associated with claudin-low tumors (Hennessy et al. 2009). On a clinical level claudin-low tumors show some chemosensitivity, whereas MBC are chemoresistant (Hennessy et al. 2009; Prat & Perou 2011). In both cases patients have a poor survival rate.

5.3.3.3. Normal-like cancers

Another subgroup that is sometimes cited as part of triple-negative cancers is normal breast-like tumors. Their gene expression levels place them in between luminal and basal-like tumors and accordingly their clinical outcome is better than that of triple-negative breast cancer, but worse than that of luminal tumors (Yersal & Barutca 2014). Many normal-breast like tumors have a ER^-, PR^-, $HER2^-$ profile, which is why they are sorted with the triple-negative tumors (Sørlie et al. 2001; Yersal & Barutca 2014). However some of them do express the estrogen and or the progesterone receptor, which is why some studies classify them as related to luminal A tumors (Prat & Perou 2011; Dai et al.

2015). Besides this ambiguous classification there are also concerns that this subtype is really an artifact due to contamination with normal breast cells (Hu et al. 2006; Yersal & Barutca 2014).

5.3.4. Molecular apocrine cancers

Historically breast tumors were analyzed for their expression of ER, PR and HER2. In 2003 a study tested two hundred cases of breast carcinoma for the presence of the androgen receptor (AR) (Moinfar et al. 2003). The authors found that in a majority of breast cancers AR was overexpressed and suggested to include the receptor into all breast cancer examinations, as it provides additional information about steroid receptors in the tumor.

Via microarray analysis a study identified a new breast cancer subgroup that is characterized by its lack of the estrogen receptor (ER^-) and its overexpression of the androgen receptor (AR^+) (Farmer et al. 2005). The group was named molecular apocrine, as their genetic expression profiles resemble that of cells found in apocrine cells. Differentiation into apocrine cells is a common pathological feature in the breast, it has been associated both with begin and malignant breast diseases, such as microscopic cysts and apocrine carcinoma (Elayat *et al.*, 2010).

Molecular apocrine cancers make up about 8-14% of breast tumors used in studies and they tend to be aggressive forms of tumors (Farmer et al. 2005; Lehmann-Che et al. 2013). While they are defined by their ER^-, PR^- and AR^+ profile, they can be both $HER2^+$ or $HER2^-$. Molecular apocrine cancers that are $HER2^+$ fall into the category of $HER2^+$ cancers and can therefore be treated with monoclonal antibodies such as trastuzumab. However it has been shown that in $HER2^+$, AR^+ it is advantageous to target the androgen receptor, as it effectively reduces cell proliferation (Ni et al. 2011). About 50-63 % of all $HER2^+$ tumors also overexpress the androgen receptor and patients could benefit from such a double treatment (Chia et al. 2015).

Molecular apocrine cancers that do not overexpress HER2, are ER^-, PR^-, $HER2^-$ and by definition belong to the group of triple-negative cancers. Between 10-53% of triple-negative cancers are AR^+ (Chia et al. 2015). The great range is due to small study cohorts and the definition of AR positivity. It was shown that triple-negative cancers that are AR^+ may benefit from a treatment with androgen inhibitors, especially in combination with PI3KCA or ERK1/2 inhibitors (Cuenca-López et al. 2014; Lehmann et al. 2014).

Luminal cancers are ER^+ and by definition have no overlap with molecular apocrine cancers. However about 84-95% or luminal cancers are AR^+(Chia et al. 2015). Historically AR expression in luminal cancers was associated with a favorable outcome, but evidence exist that AR may act as an oncogene in the case of tamoxifen-resistance (Chia et al. 2015). In fact overexpression of the androgen receptor has even been observed to induce tamoxifen-resistance (De Amicis et al. 2010).

The development of microarray studies has allowed researchers to analyze large numbers of genes expressed in different tumor samples. The expression levels have led to the identification of groups and subgroups, the most important are named above. However other subgroups exist and the above mentioned classes can be further subdivided. Also it has to be noted that these groups are not mutually exclusive. For example, a cancer can be categorized at the same time as triple-negative, claudin-low and molecular apocrine, depending on the markers used.

In a clinically setting only groups are relevant that will lead to a form of targeted treatment, such as the hormone therapy for luminal cancers or the monoclonal antibodies for $HER2^+$ cancers. Maybe with the advent of anti-AR based medications, the androgen receptor positive groups will gain in importance. But even the breast cancer subtypes, for which a targeted treatment exists, show resistance to treatment in some cases, but not in others. Therefore, more markers are needed that will correctly identify the response of a cancer to a treatment.

5.4. Breast cancer stem cells

The biggest challenges in cancer treatment are the fight against relapse and metastasis. Early on cancer researchers had noticed that one single tumor was formed of a very heterogeneous set of cells (Heppner et al. 1983). Debate arose around the question how this tumor heterogeneity occurred. Some argued that some tumor cells underwent sporadic mutations, which were then propagated in a clonal manner. In this model every cell has the potential to form a new tumor, which will then again become heterogeneous through mutations (figure 17).

An opposing model stated that only a small number of cells had an unlimited capacity to self-renew and form a tumor colony, whereas the majority of tumor cells had a finite ability to proliferate. The

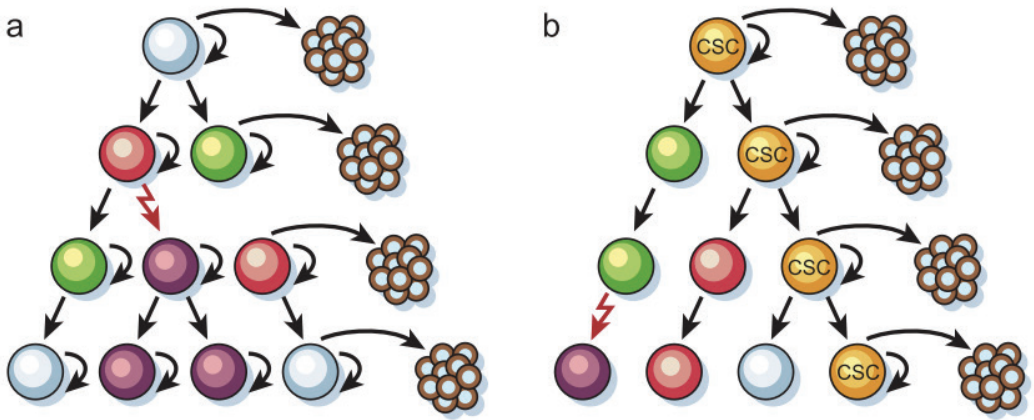

Figure 17: Two models of heterogeneity in tumors. (a) The heterogeneity in tumors is caused by sporadic mutations. Many of the different tumor cells have the ability to proliferate and form new tumors. (b) Only cancer stem cells (CSC) (yellow) have the capacity to proliferate and form new tumors. They can also differentiate into many different cells types, which in turn might be altered by mutations.

Image from Reya *et al.*, 2001

term tumor stem cell was coined (Bergsagel & Valeriote 1968). In 1994 a study demonstrated that specific subpopulations of leukemia cells had enriched tumor forming potential when xenografted into mice, whereas other cells lacked this ability (Lapidot et al. 1994).

5.4.1. Breast cancer stem cell markers

In 2003 Al-Hajj and colleagues sorted cells according to cell surface markers CD44 and CD24 as well as to certain lineage markers. They showed that as little as 100 cells with a $CD44^{+}/CD24^{-/low}/Lin^{-}$ phenotype were able to form tumors in mice, whereas tens of thousands of cells with alternative phenotypes failed to do so (Al-Hajj et al. 2003). Furthermore, the tumors generated by $CD44^{+}/CD24^{-/low}/Lin^{-}$ could be serially passaged. The newly generated tumors would contain $CD44^{+}/CD24^{-/low}/Lin^{-}$ as well as the phenotypically diverse subpopulations found in the original tumor. Therefore, the cells fulfilled all characteristics of cancer stem cells:

1) tumor formation *via* xenografts in mice
2) self-renewal in secondary mice
3) "differentiation" into cells with non-stem cell characteristics (Mcdermott & Wicha 2010).

Following the work from Al-Hajj et al (2003), a number of other breast cancer stem cell markers were identified, among them the CD49f (Cariati et al. 2008), CD61 (Vaillant et al. 2008), CD133 (Wright et al. 2008) and the Aldehyde Dehydrogenase 1 (ALDH1) (Ginestier et al. 2007). But not all markers are expressed in all types of tumors and it is possible that each cancer subtype has its own set of markers (Owens & Naylor 2013).

The question as to how well the different markers can predict clinical outcome has not yet been answered in full. One study found that the markers $CD44^{+}/CD24^{-}$ alone did not predict overall survival (Bane et al. 2013). However the combination of several markers allowed to identify high risk patients in breast cancer (Neumeister et al. 2010). Another study that analyzed the data of 12 different studies came to the conclusion that breast cancers with a high proportion of cancer stem cells (CSC) are correlated with poor outcome (Zhou et al. 2010).

5.4.2. Stem cellness and therapy resistance

It has to be noted that cancer stem cells are not a fixed set of cells. In fact it has been shown that CSC shift between a stem-like and a non-stem-like state (Meyer et al. 2010; Gupta et al. 2011; Wang et al. 2016). This kind of plasticity raises the question of how stem-cellness is acquired and maintained. Signaling pathways Hedgehog, Notch and Wnt have been linked to stem-cell self-renewal and differentiation (Bozorgi *et al.*, 2015; Paula and Lopes, 2017). Furthermore, the tumor

microenvironment known as *niche* plays an important role in maintaining CSC (Kise *et al.*, 2016). Factors such as hypoxia, elevated levels of cytokines as well as paracrine and autocrine signaling influence the survival of stem cells (Yang et al. 2017).

Several mechanisms have been identified that convey therapy resistance to cancer stem cells. It was found that some CSC overexpress ATP-binding cassette (ABC) transporters (Moitra & Karobi 2015). These transporters are able to transport cytotoxic drugs out of the cell and thereby help cells evade cell death (Leonard *et al.*, 2003). Many cancer therapies induce DNA breaks, failure to repair these breaks then lead to apoptosis. CSC can avoid cell death through enhanced DNA repair mechanisms (Bao *et al.*, 2006; Phillips *et al.*, 2006; Peitzsch *et al.*, 2013).

In response to radiation therapy cells produce reactive oxygen species (ROS). Excess amounts of ROS that will interact with DNA, proteins or lipids can induce apoptosis (Cook et al. 2004). Enzymes involved in ROS scavenging, that is the elimination of excess ROS, were found to be overexpressed in cancer stem cells (Bozorgi *et al.*, 2015). One of the enzymes helping CSC fight reactive oxygen species is the aldehyde dehydroxygenause (ALDH). Overexpressed in many cancer stem cells (Ginestier et al. 2007) it fights ROS directly by removing oxygen radicals and indirectly by removing antioxidant compounds (Singh et al. 2013).

5.4.3. Epithelial-to-mesenchymal transition

The transition from epithelial to mesenchymal cells is a fundamental process that is necessary for development, wound healing and tissue regeneration. Epithelial cells are characterized by tight cell-cell junctions and a distinct apical versus basolateral polarity. Typically, epithelial cells form a barrier between two compartments. Mesenchymal cells on the other hand have a connecting or scaffolding role. They lack the tight cell-cell junctions and the polarity found in epithelial cells. EMT can be induced *via* different pathways, which will trigger a cascade of signaling events (figure 18). Some of the genetic changes during EMT include the upregulation of transcriptional repressors SNAIL, SLUG, TWIST and ZEB1. They will in turn downregulate the expression levels of the adherens junction protein E-cadherin. Reduced E-cadherin expression leads to the collapse of adherens junctions. Cell polarity is lost through modulation of Rho GTPase function. These changes lead to a breakdown of the inter-cell connections and to a gain in motility of the cells (Singh & Settleman 2010). During EMT cells stop being part of a bigger ensemble, tightly communicating *via* cell-cell junctions. Instead they become individuals with increased invasive characteristics.

While EMT is a natural process during development or wound healing, it has also been linked to cancer stem cells (Singh & Settleman 2010). It was shown that $CD44^{low}/CD24^{+}$ could be transformed into breast cancer stem cells ($CD44^{+}/CD24^{-}$), by inducing EMT (Morel et al. 2008).

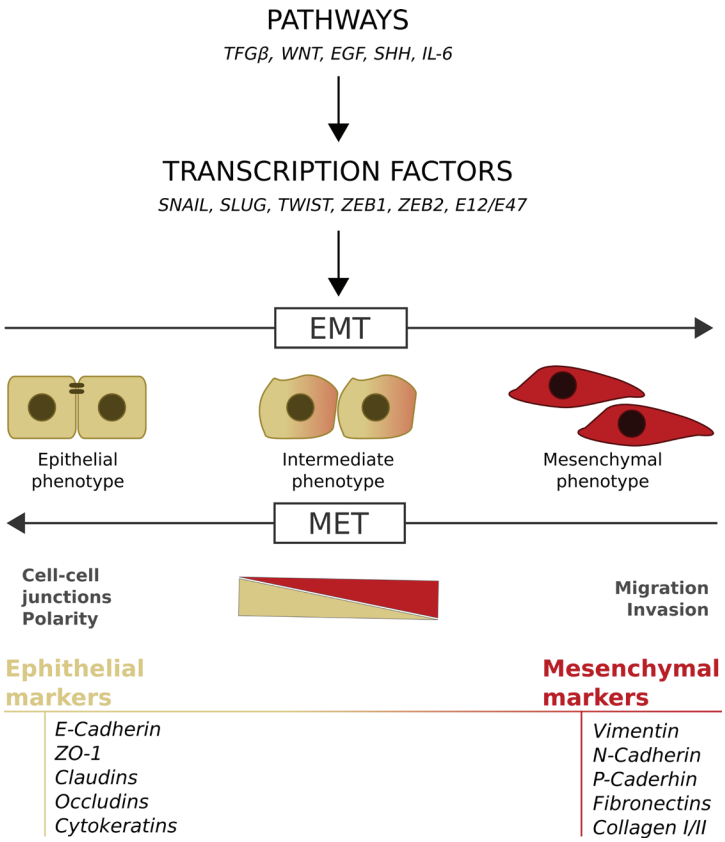

Figure 18: The Epithelial-to-mesenchymal-transition (EMT). Different pathways can induce EMT. Transcription factors such as SNAIL, SLUG or TWIST will be activated and cells transform from an epithelial to a mesenchymal phenotype. While epithelial cells are tightly interconnected and have apical-basal polarity, mesenchymal cells are detached and show high capacity to migrate and invade other tissue. The transition is a continuous process in which epithelial markers are progressively diminished and mesenchymal markers are gained. The reverse process is called mesenchymal epithelial transition (MET)

Furthermore, the E-cadherin repressor TWIST is able to promote cancer stem cellness (Liang et al. 2015).

It is believed that metastasis arise from cancer stem cells that underwent EMT. These cells have lost their connections with the other tumor cells and are thus able to detach from the main tumor. They will first invade the extracellular matrix and might eventually enter a blood vessel. With the blood flow the cell is transported to distant parts of the body, where the cell will exit the blood vessel. If the cell escapes the immune system, it can now settle in the new site. Here it will undergo a mesenchymal-epithelial transition (MET) and start a new tumor (figure 19).

In line with this theory, two types of stem cells were identified: EMT-CSC and MET-CSC (Liu et al. 2014). The former have a $CD44^{-}/CD24^{-}$ signature, low levels of epithelial marker E-cadherin and high levels of mesenchymal marker vimentin. The latter are characterized by an $ALDH^{+}$ phenotype, high levels of E-cadherin and low levels of vimentin. Mesenchymal cancer stem cells (EMT-CSC) were found to be localized at the tumor invasive front, from where they can detach and colonize other body parts. Epithelial cancer stem cells however (MET-CSC), localize more centrally, which is in line with the idea of MET-CSC being the founding cells of secondary tumors (Liu et al. 2014).

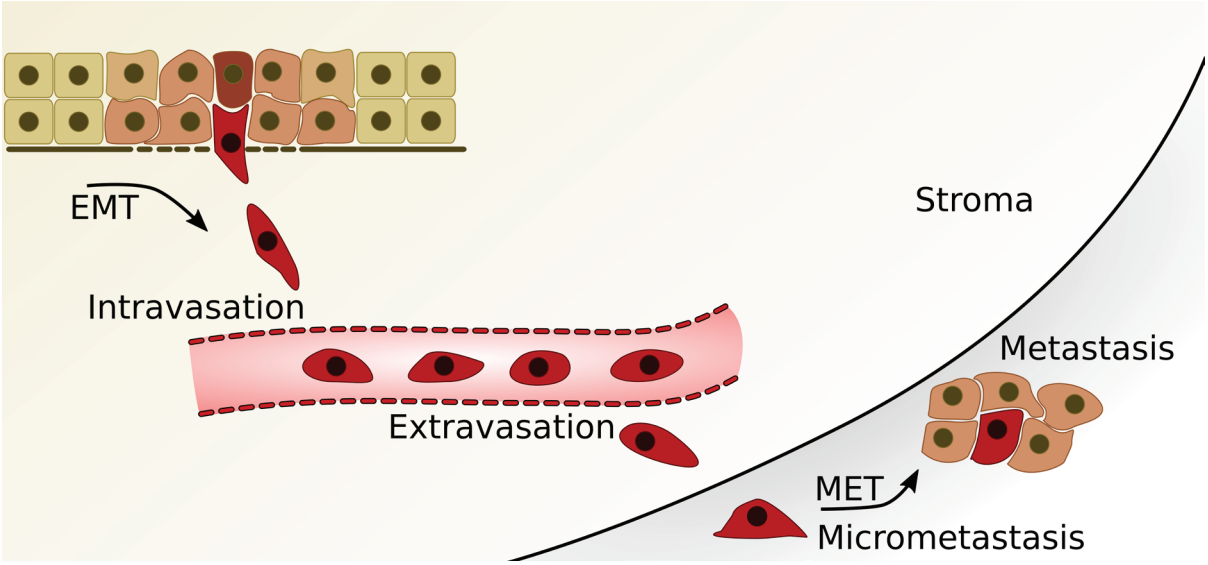

Figure 19: The role of EMT and MET in the formation of metastasis. At the primary tumor site cells undergo EMT, which makes them mobile. They enter the blood stream and get transported to a distant site, where they exit the bloodstream. First the tumor cells settle as micrometastasis, if they escape the immunsystem they can undergo MET, divide and form larger metastasis.

Given the role of cancer stem cells in therapy resistance and metastasis, new drugs have to be developed that target specifically these tumor initiating cells. One of the most promising compounds is metformin, a drug normally used to fight diabetes. It was shown that metformin targets specifically $CD44^{-}/CD24^{-/low}$ cells and that this is sufficient to overcome trastuzumab resistance in $HER2^{+}$ tumors (Cufi et al. 2012). Furthermore, in combination with the chemotherapy drug doxorubicin, metformin selectively eradicates cancer stem cells in breast cancer (Hirsch et al. 2009). A number of other cancer stem cell inhibitors are currently being tested as potential drug targets (Paula & Lopes 2017), but more research is needed to understand the complexity of cancer stem cell regulations.

6. Purpose of this study

The aim of this study is to further the understanding of the role played by RPC32α tumor development. As a first step an appropriate tumor model system needs to be established and validated for its clinical relevance. Once a model system has successfully been implemented in the laboratory, the function of RPC32α is to be analyzed by molecular and biochemical techniques. Any results obtained *in vitro* will need to be validated in an *in vivo* model.

While the elucidation of RPC32α is the main goal of this study, the mission is too multifaceted, to be accomplished in only three years time. Consequently, this work is meant to lay to be the foundation for future studies. This will be done on the one hand by identifying cellular processes in which RPC32α is involved and by creating valuable tools and techniques to be used by those who follow this work.

Chapter II

Materials & Methods

Materials

1. Cell lines and culture conditions

All cell lines used are human epithelial cells that were cultured at 37° C avec 5% of CO_2. Unless marked otherwise, cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum, 100 units/mL of penicillin and 100 µg/mL of streptomycin.

The following cell lines were used:

MCF-10A (ATTC CRL-10317): This cell line is derived from human fibrocystic mammary tissue. It is non tumorigenic and immortalized spontaneously (Soule *et al.*, 1990). The cells were cultured in Dulbecco's Modified Eagle's Medium/Nutrient F12 (DMEM/F12) supplemented with 10% horse serum, 100 units/mL of penicillin, 100 µg/mL of streptomycin, 0.01 mg/mL insulin, 20 ng/mL of epidermal growth factor as well as 0.5 mg/mL de hydrocortisone .

BT-474 (ATTC HTB-20): This luminal cell line was isolated from solid, invasive ductal breast carcinomas (Lasfargues *et al.*. 1978). It is aneuploide with a chromosome count in the hypertetraploid range. It was cultured in Roswell Park Memorial Institute medium (RPMI) 1640, supplemented with 10% fetal bovine serum, 100 units/mL of penicillin and 100 µg/mL of streptomycin .

MCF7 (ATTC HTB-22): Originally this cell line was taken from a pleural effusion of a patient with metastatic breast carcinoma (Soule *et al.*, 1973). The cells have a chromosome count that is hypertriploid to hypotetraploid.

MDA-MB231 (ATTC CRM-HTB-26): First established in 1973 this cell line was obtained from a pleural effusion of a breast carcinoma patient (Cailleau *et al.*, 1974). It has a chromosome count in the near triploid range.

BT-549 (ATCC HTB-122): The original cells were isolated from a primary breast tumor in 1978 by Curtinho and Lasfargues (Neve *et al.*, 2006). The cells have a chromosome count in the hypertetraploid range.

MDA-MB468 (ATCC HTP-132): Taken from a pleural effusion of an adenocarcinoma, this cell line was established in 1978 (Cailleau, Olivé and Cruciger, 1978). Chromosome counts are in the hypotriploid range.

MDA-MB453 (ATCC HTP-131): This cell line was established from a pericardial effusion (Cailleau, Olivé and Cruciger, 1978). It presents a chromosome count in the hypo- to near tetraploid range.

As part of this thesis project several genetically modified cell lines were created.

MDA-MB231-RPC32α-KO cell lines: The mother cell lines MDA-MB231 was altered using CRISPR-Cas9. The gene POLR3G, which codes for RPC32α, was cut directly downstream of the start codon and the double stranded cut was repaired *via* non-homologous end joining (NHEJ). Four RPC32α knock-out clones were amplified and used for further experiments. Clones 1, 2 and 4 suffered a deletion of a guanine downstream of the start codon, clone 3 is missing a thymine and guanine immediately behind the start codon.

MDA-MB231-luciferase: A gene coding for the luciferase has been inserted into the MDA-MB231 genome *via* transduction. These cells were used for *in vivo* experiments in mice. Injection of luciferine into the mice made the cells luminescent which made it possible to follow cell growth *in vivo*.

MDA-MB231-RPC32α-KO-luciferase: The RPC32α clone 1 had been transduced with a gene coding for the luciferase protein. This cell line was used for *in vivo* mice experiments.

All genetically modified cell lines were cultured under the same conditions as the mother cell line.

2. Bacterial strains

XL1-blue: This *E.coli* strain has a natural resistance against tetracycline and it has an *endA1 gyrA96 (nal^R) thi-1 recA1 relA1 lac glnV44 F'[:: Tn10 proAB$^+$ lacI^q Δ(lacZ) M15] sdR17 (r_K^- m_K^+)* genotype. This strain was used to amplify plasmids.

It is cultured in Lysogeny broth (LB) medium (10 g tryptone, 5 g yeast extract, 10 g NaCl) completed with Ampicillin (100 µg/mL). Liquid cultures were kept at 37° C under orbital agitation (200 RPM). For solid cultures agar (20 g/L) was added to the LB medium and plates were grown at 37° C.

3. Plasmids and expression vectors

pSpCas9(BB)-2A-Puro (PX459) (Addgene): This plasmid was used for all CRISPR-Cas9 constructs. It contains the gene coding for the nuclease Cas9, tagged with a triple-Flag tag and is preceded by a chicken-β-actin promotor. Furthermore, it contains the guide RNA scaffold, in front of which can be cloned the desired target sequence. The whole being under the control of a U6 promotor. For the selection in mammalian cells and bacteria the plasmid has respectively a puromycine and an ampicillin resistance gene. Using the px459 plasmid as a backbone, two different CRISPR plasmids have been created. In both cases the target sequence was integrated into the plasmid by simultaneous digestion and ligation using the *Bbs*I restriction enzyme and the T4 ligase.

CRISPR1: The target sequence GTATAACTGGTTCTGATGGCT guides the Cas9 to cut immediately after the start codon.

CRISPR2: The target sequence GGGTGGTTTCAACACTACATC directs the Cas9 to cut 84 nucleotides downstream of the start codon.

4. Primers and oligonucleotides

4.1. Primers for RT-qPCR

The list of primers used for RT-qPCR:

Target	Forward	Reverse
ESR1 (ER)	CTCTATGACCTGCTGCTG	CTTTGGTCCGTCTCCTC
AREG (AR)	CAAGTCACACATGGTGAGCGT	TCTTGGGCACTTGCACAGAG
ERBB2 (HER2)	GCCAGTGTGAACCAGAA	CTCTTGATGCCAGCAGAA
FOXA1 (FOXA1)	CGGAGCAGCAGCATAAG	GCAACGTAGAGCCGTAAG
POLR3G (RPC32α) exon 1	CGTTCTCTGCCGTCACCC	AAAGGCACTGCTCCCTAAGTCTC
POLR3G (RPC32α) exon 6	CGCAGGCAAAGGCACAC	CCTCTTTTTTCCAATTCCTCCA
POLR3GL (RPC32β)	CCAAGAGAGATGTGGAGCGTTATT	TCCAATCGATGGCATTGTCA
POLR3C (RPC62)	ACTGGTGCAGAGGAAGCACA	TCTAGCTGCTGACGTTCAGGAG
POLR3F (RPC39)	AGAAGGCACAGTTGGCAGTGT	TGGGAGGGATGATTGGATTG
POLR3D (RPC53)	ACCCTGGCTGACCTGACAGA	AGGAGTTGCACCCTTCCAGA
rRNA 5S	CTGAACGCGCCCGATCT	GCGGTCTCCCATCCAAGTAC
tRNA^{METi}	AGAGTGGCGCAGCGGAA	TAGCAGAGGATGGTTTCGATCC
tRNA^{METe}	GCCTCGTTAGCGCAGTAGGTA	GAGGATCGAACTCACGACCTTC
RNA 7SK	TCTTCGGTCAAGGGTATACGAGTAG	CAAATGGACCTTGAGAGCTTGTT
RNA vault1	GGCTGGCTTTAGCTCAGCG	TCTCGAACAACCCAGACAGGT
RNA U6	CTCGCTTCGGCAGCACATATA	AACGCTTCACGAATTTGCG
BC200	GGTGGCTCACGCCTGTAATC	GAACTCCTGGGCTCAAGCTATC
MRP RNA	GCTTCCCACTCCAAAGTC	CGTAACTAGAGGGAGCTGA

4.2. Primers for PCR

List of primers used for PCR

Purpose	Name	Orientation	Sequence
Testing for CRISPR1 mutations	CRISPR1 test a	Forward	GGGGTGCAGTTTTATATTTTGCT
	CRISPR1 test b	Forward	GTCAACTGTGCTACTTAAGGG
	CRISPR1 test a	Reverse	TGCTAAATCCAACAGCCTCAA
	CRISPR1 test b	Reverse	GGGTGGTTTCAACACTACATC
Testing for CRISPR2 mutations	CRISPR2 test a	Forward	GGGAATAAAGGAAGAGGACGTG
	CRISPR2 test a	Reverse	CTCTGAAAAATGAAAAAACACACTATG

4.3. Oligonucleotides

List of oligos used to test for G-quadruplex structures

Number	Sequence
1	TGGGAGGGGTGGTGAGGGTAGAGGGGAATTAAAGGGG
2	TGGGAGGGGTGGTGAGGGTAGAGGGGA
3	AGGGGTGGTGAGGGTAGAGGGGAATTAAAGGGG
4	TGGGAGGGGTGGTGAGGGTAGAGGTTAATTAAAGGGG
5	TGGGAGGGGTGGTGAGGGTAGAGTTGAATTAAAGGGG
6	TGGGAGGGGTGGTGAGGGTAGATTTAAATTAAAGGGG
7	TGGGATGGGTGGTGATTGTAGATTTAAATTAAAGGGG
8	TGGGAGGGGTGGTGAGGGTAGATTTAAATTAAATTTG
9	TAGGATGTTTAATGAGGGTAGATTTAAATTAAAGGGG
10	TGGGATTTATGGTGAAAATAGATTTAAATTAAAGGGG
11	TAGGATGTTTATTGAAAATAGATTTAAATTAAATTTG
12	TTTTATTTATATTGAAAATAGATTTAAATTAAATTTG

5. Antibodies

List of antibodies used in this study

Target	Reference number	Supplier
RPC32α	SC-21754	Santa Cruz
RPC32β	HPA027288	Sigma-Aldrich
RPC62	monoclonal antibody produced in the lab of Robert Roeder, Rockefeller University, NY, USA	
Flag-tag	200470-21	Agilent
β-actin	SC-81178	Santa Cruz

Methods

1. Creation of competent bacteria

The bacterial preculture of 10 mL in Lysogeny broth (LB) medium completed with Tetracycline (20 µg/mL) was grown overnight at 37° C. The next day 1 L of bacterial culture was inoculated with the

preculture and left to grow at 37° C under agitation (200 RPM). When the exponential growth phase was reached (OD_{600nm} between 0.6 and 0.8), the culture was stopped and the bacteria were centrifuged (4000 RPM, 15 min, 4°C) and washed with sterile water three times. After the final wash step the pellet was resuspended in 10% glycerol, aliquoted and stocked at -80° C.

2. Bacterial transformation

50 µL competent bacteria were mixed with 5-10 ng DNA. The transformation was performed via an electric pulse (1800 V; 1,5 msec). Following the transformation, the bacteria were resuspended in 200 µL liquid LB medium. Depending on the efficiency of the transfection 10 – 200 µL were given on an LB agar plate and left to grow at 37° C.

3. Plasmid extraction

3.1. Miniprep

A 5 mL culture of LB medium plus Ampicillin (100 µg/mL) was left to grow overnight (37° C, 200 RPM). The next day the bacteria were pelleted (4000 RPM, 15 min, 4° C) and resuspended in 100 µL of solution I (EDTA 10 mM, Tris-HCl 25 mM pH 8, RNase (DNase free) 50 µg/mL). The resuspended bacteria were transferred in to a 1.5 mL reaction tube.

200 µL of solution II (NaOH 0.2 M, SDS 1 %) were added and the contents was carefully mixed by inverting the tubes 3-4 times. The mix was incubated for 5 minutes at room temperature. To stop the reaction 150 µL of ice cold solution III (potassium acetate 3M pH 4.8) were added and again the tubes were inverted carefully 3-4 times. The debris was pelleted (12 000 RPM, 25 minutes, 4°C) and the supernatant was transferred to a new 1.5 mL tube.

The plasmid DNA was pelleted with 1 mL ice cold ethanol 100% (12 000 RPM, 30 minutes, 4°C). Subsequently the DNA pellet was washed with 70 % ethanol (12 000 RPM, 5 minutes, 4° C), dried (37° C, 5 min) and resuspended in 30 µL sterile water.

Plasmid concentrations were measured using a Nanodrop 1000 (ThermoScientific). All plasmids were verified by sequencing before utilization.

3.2. Midiprep

A 5 mL preculture of LB medium plus Ampicillin (100 µg/mL) was left to grow overnight (37° C, 200 RPM). The following day a 100 mL culture was inoculated with the preculture and again left to grow overnight. In the morning the bacterial culture was divided into two 50 mL Falcons and centrifuged (4000 RPM, 15 min, 4° C). The pellet was resuspended in 2 mL solution I (EDTA 10 mM, Tris-HCl 25 mM pH 8, RNase (DNase free) 50 µg/mL). 2 mL of solution II (NaOH 0.2 M, SDS 1 %) were added to the resuspended pellet and the whole was mixed gently by inverting the tubes 3-4 times. The mix was left to incubate on ice for 5-10 minutes.

To stop the reaction 2 mL of solution III (potassium acetate 3M pH 4.8) were added. The tubes were gently inverted 3-4 times and left to incubate on ice for 5 minutes, before the debris was pelleted (4000 RPM, 15 minutes, 4° C). The supernatant was transferred to a new falcon tube 50 mL and the DNA was precipitated with 6 mL isopropanol (4000, 30 minutes, 4° C).

The pellet was resuspended in 500 µL TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) and treated with 1.5 µL RNase (10 µg/µL; 37° C, 15 minutes) and 2.5 µL proteinase K (20 µg/µL; 37° C, 15 minutes). Subsequently 1 volume of a phenol-chloroform-isoamyl alcohol mixture (25:24:1, v/v) was added and the phases were separated by centrifugation (10 000 rpm, 5 minutes, room temperature).

The aqueous phase was transferred to a new tube and the DNA was pelleted with 2 volumes of 100 % ethanol and 10% NaCl 5M. The mix was left at least 30 minutes at -20° C before the DNA was pelleted (140 000 RPM, 30 minutes, 4° C). The two pellets were reunited in 1000 µL of TE buffer and precipitated with 400 µL of PEG buffer (PEG 30% (8000), NaCL 1.6M). The reaction was left overnight at 4° C.

The following day the DNA was pelleted (10 000 RPM, 30 minutes, 4°C), washed with 70% ethanol (10 000 RPM, 10 minutes, 4° C), dried (5 minutes, 37° C) and resuspended in 100 µL of sterile water.

4. CRISPR

4.1. Design

The sequence for the guide-RNA was identified using either the website chop-chop (https://chopchop.rc.fas.harvard.edu) or CRISPR Design (http://crispr.mit.edu:8079).

The complement strand of the guide-RNA was created and the necessary restriction sites were added. In the case of the plasmid px459 a *Bbs*I site was used, therefore a CACCG overhang was added to the 5' site of the upper strand and a CAA overhang was added to the 3' end of the lower strand. A

single C nucleotide was added to the 5' end of the lower strand for more stability and the two sequences were ordered from Sigma-Aldrich.

Upper strand: 5' – **CACCG**NNNNNNNNNN – 3
Lower strand: 3' – **C**NNNNNNNNNN**CAAA** – 5'

For this study two different guide RNAs were used:

Guide RNA 1:

Upper strand: 5'-**CACCG**TATAACTGGTTCTGATGGCT-3'
Lower strand: 5'-**AAAC**AGCCATCAGAACCAGTTATA**C**-3'

Guide RNA 2:

Upper strand: 5'-**CACCG**GGTGGTTTCAACACTACATC-3'
Lower strand: 5'-**AAAC**GATGTAGTGTTGAAACCACC**C**-3'

4.2. Annealing and phosphorylation

To create a double stranded DNA sequence that could be inserted to the plasmid, the two oligos had to be annealed and phosphorylated. For this, the following mix was prepared:

The reaction was incubated in a thermocycler:

1 µL	oligo upper (100 µM)
1 µL	oligo lower (100 µM)
1 µL	10x T4 Ligation Buffer (NEB)
6.5 µL	H_2O
0.5 µL	T4 PNK (NEB)
Total volume 10 µL	

37° C for 30 minutes
95° C for 5° C
Descend to 25° C with 5°C per minutes
The final reaction was diluted 1:10 for further use.

4.3. Digestion and ligation

The vector was then digested and ligated with the guide-RNA in one single step.

1 µL	vector px459 (50 ng)
1 µL	oligos phosphorylated and hybridized diluted 1 : 10
2 µL	10 x T4 Ligation Buffer
1 µL	*Bbs*I restriction enzyme (NEB)
1 µL	T4 ligase (NEB)
4 µL	H_2O
Total volume 10 µL	

The reaction was incubated in a thermocycler with 20 cycles of 37° C for 5 minutes and 23° C for 5 minutes. The plasmid was amplified in XL1-blue *E.coli* bacteria.

4.4. Cellular transfection

MDA-MB231 were grown on 10 cm diameter culture dishes to 70% confluency. To prepare the transfection 1000 µL Opti-Mem (Gibco), 6 µg of CRISPR plasmid DNA and 6 µL Plus-Reagent were mixed and incubated for 5 minutes at room temperature. Following incubation 15 µL of Lipofectamine LTX (Invitrogen) were added. The mix was vortexed vigorously and left to incubate for 30 minutes at room temperature. 9 mL fresh culture medium were given to the cells before the DNA/Lipofectamine Mix was added. The cells were left to grow for 48 hours at 37° C.
To select cells with a plasmid, puromycin was added to the culture medium (2 µg/mL) for three days. After selection, cells were trypsinized and suspended in 1 mL cell culture medium. The suspension was serially diluted (v/v) to 1:10 000; 1:50 000; 1:100 000 and 1:500 000. Each dilution was plated into a 96 well plate with 100 µL per well. The last dilution that still yielded colonies was used for further testing. Cells were amplified until they could be tested for potential frame shift mutations that would lead to a knock-out.

4.5. Identification of potential knock-out cell lines

The genomic DNA of the transfected cells was extracted. The target zone of CRISPR1 or CRISPR2 was amplified *via* PCR using combinations of the primer pairs *CRISPR1 test a* or *b* and *CRISPR test 2* respectively. Of the PCR product 1-2 µL were given on to a 10% acrylamide gel (12,5 mL acrylamide

40% 19:1; 2.5 mL TBE 10x (Euromedex), 500 µL APS, 50 µL TEMED and 34.45 mL H_2O). The gel was left to run overnight (250 V, 4° C). The gel was revealed using syber gold (Sigma Aldrich) and the Gel Doc XR+ Imager (BioRad). Lanes that showed several bands were excluded, as it could not be a homozygous mutant. Single bands that were slightly higher or lower than the wild type band were selected for further analysis. The identified clones were again used for a PCR. The PCR product was sent to sequencing by GATC.

5. DNA extraction

5.1. DNA extraction from cell culture dishes

For standard DNA either a 15, 10 or 6 cm diameter culture dish was used that had grown to confluency. The cells were trypsinized and pelleted (1000 RPM, 5 minutes, RT). The pellet was washed with 2.5 volumes of solution A+ ($MgCl_2$ 1.5 mM; KCl 10 mM; Tris-HCl 20 mM; NP40 0,1%) and centrifuged (4500 RPM, 15 minutes, 4°C). The supernatant was eliminated and the pellet was washed with solution A ($MgCl_2$ 1,5 mM; KCl 10 mM; Tris-HCl 20 mM) and centrifuged (4500 RPM, 15 minutes, 4°C). Again the supernatant was eliminated, the pellet was resuspended in 2.5 volumes of cell lysis buffer (Tris-HCl pH 7.5 10 mM, EDTA 10 mM, NaCl 10 mM, SDS 0,5%) and the RNase was added (20 µg/mL). The reaction was incubated at 37° C for one hour, then the proteinase K was added (2 µg/mL). The mix was left to incubate at 55° C for least three hours, but preferably overnight.
If necessary, the total volume was filled up to 300 µL and a phenol/chloroform extraction was performed. One volume phenol/chloroform mix pH 8 was added to the sample. After vigorous vortexing the sample was centrifuged (10 000 RPM, 5 minutes, RT) and washed with one volume of chloroform (10 000 RPM, 5 Minutes, RT). The supernatant was precipitated with NaAc ($1/10^{th}$ of the volume of the supernatant, Sodium Acetate 3M pH 5,2) and 100% ethanol (2.5 volumes of the supernatant). The reaction was left at least 30 minutes at -80° C. The DNA was pelleted (14 000 RPM, 30 minutes, 4° C), washed with 1 mL 70% ethanol (14 000 RPM, 5 minutes, 4° C) and dried (5 minutes at 37° C) and resuspended in 50, 30 or 20 µL sterile water for 15, 10 or 6 cm culture dishes.

5.2. DNA extraction from 96 well plates

In a 96 well plate the cells were grown to confluency. The cells were washed twice with PBS, before adding 50 µL of lysis buffer (10 mM Tris-HCl pH 7.5; 10 mM EDTA; 10 mM NaCl; 0,5% SDS, proteinase

K (1 mg/mL)). The plate was closed with parafilm, wrapped in moist tissue, sealed in a bag and placed at 55° C overnight.

The next day 100 µL of ice cold 100% EtOH with NaCl 0.2 M was added to each well and the plate was left to incubate at room temperature for 30-60 minutes. The plate was then overturned to empty it of the ethanol and the wells were washed twice with EtOH 70%. Finally, the plate was left to dry on the bench for 20 minutes. The DNA was resuspended in 20 µL of sterile water. The plate was place at 4° C overnight before the DNA was used for further experiments.

6. RNA extraction

Cells from a 15 cm diameter culture dish were trypsinized and pelleted (1000 RPM, 5 minutes, RT). The pellet was resuspended in 1 mL TRI reagent (Molecular Research Center, Inc.) and transferred to a 2 mL reaction tube containing 25 µL of glycogen. After vortexing, 200 µL of chloroform were added and the whole was shaken vigorously for 15 seconds. Afterwards the mix was left on ice to incubate for 15 minutes. Phases were separated by centrifugation (10 000 RPM, 10 minutes, 4° C) and the aqueous phase was transferred to a new tube. One volume of TRI reagent (Molecular Research Center, Inc.) and 1/5 volume of chloroform were added to the mix. Again the tube was shaken vigorously for 15 seconds and left to incubate on ice for 10 minutes. Phases were separated by centrifugation (10 000 RPM, 10 minutes, 4° C). The aqueous phase was transferred to a new reaction tube and mixed with one volume of chloroform. The mix was shaken vigorously for 15 seconds and left to incubate on ice for 5 minutes. The phases were separated by centrifugation (10 000 RPM, 5 minutes, 4° C) and the aqueous phase was transferred to a new tube. The RNA was precipitated with one volume of isopropanol. After mixing well, the tube was left for 10 minutes at room temperature, before being left overnight at -20° C.

The next morning the RNA was pelleted (13 000 RPM, 30 minutes, 4° C), washed with 75% ethanol (7 500 RPM, 10 minutes, 4° C) and dried (5 minutes, 37° C). The pellet was resuspended in 50 µL RNase free sterile water. RNA concentration was measured using the Nanodrop 1000 (Thermo Fisher Scientific).

7. RNA quality control

All RNA extracts were quality tested on a 2100 Bioanalyzer (Agilent) according to the manufacturer's instructions. Only RNAs with a RIN above 9 were kept for further analysis.

8. Reverse transcription

For each reaction 2 µg of RNA were used. To the reaction were added 10 µL of random primer (15 ng/µL) (Thermo Fishern Scientific) and 1 µL of oligo d(T) Primer (150 ng/µL) (Thermo Fisher Scientific). The mix was completed up to 20 µL with sterile water. In a thermocycler the reaction was incubated at 65° C for 5 minutes to unfold any secondary structures and to allow the primers to anneal to the RNA.

To each tube was added a reverse transcription mix (8 µL 5x Buffer; 2 µL dNTPs (10 mM), 1 µL Ribolock RNase Inhibitor (Thermo Fisher Scientific), 1 µL Maxima Reverse Transcriptase 200 U/µL (Thermo Fisher Scientific) and 8 µL H_2O). The reverse transcription was performed using a thermocycler (10 minutes, 25° C; 1 hour, 50° C and 15 minutes 70°C). Finally 60 µL of sterile water were added to each tube to obtain a final cDNA concentration of 20 ng/µL.

9. RT-qPCR

All RT-qPCR reactions were done using a CFX96 Touch Real-Time PCR Detection System (BioRad). For standard reactions 4 ng of cDNA were used. To this were added 5 µL Sso Advanced universal SYBR Green Supermix (BioRad), 3 µL mix of forward and reverse primers (1.6 µM) and sterile water to give a total reaction volume of 25 µL. On every plate each gene was always tested in replicate.

The standard protocol was 95° C for 30 seconds, followed by 40 cycles of 95° C for 5 seconds and 60° C for 10 seconds. For the melt curve the temperature rose from 65° C to 95° C with an increment of 0.5° C every 5 seconds. The reaction finished with 1 minute at 22° C.

The qPCR data was analyzed using the program CFX Manager 3.0 (BioRad). The average cycle threshold (CT) values of the two replicates of each gene was calculated. The CT of each gene tested is corrected by the average CT of the two housekeeping genes:

$$CT_{\text{gene tested}} - CT_{\text{housekeeping gene}} = \Delta CT.$$

This new ΔCT was normalized with the ΔCT of the control line:

$$\Delta CT_{\text{breast cancer line}} - CT_{\text{control}} = \Delta\Delta CT$$

The relative quantity of a gene was calculated as fold change using the formula:

$$FC = 2^{-\Delta\Delta CT}.$$

Each RT-qPCR was done in triplicate using biological replicates.

10. Protein extraction

For all protein extractions a 15 cm diameter cell culture dish was used that had grown to confluency. The cells were trypsinized, pelleted (1000 RPM, 5 minutes) and resuspended in two volumes of solution A+ ($MgCl_2$ 1,5 mM; KCl 10 mM; Tris-HCl 20 mM; NP40 0,1%). This solution would break the cell membrane, after centrifugation (5 000 RPM, 15 minutes, 4° C) the supernatant containing the cytoplasmic proteins was transferred to a new tube and supplemented with 30% of glycerol. The pellet was resuspended in 1 mL of solution A ($MgCl_2$ 1.5 mM; KCl 10 mM; Tris-HCl 20 mM) and centrifuged again (5 000 RPM, 15 minutes, 4° C). The pellet was resuspended in one volume of solution B (glycerol 50%, $MgCl_2$ 1,5 mM; KCl 10 mM; Tris-HCl 20 mM) and supplemented with one third of the total volume of solution C (glycerol 50%, $MgCl_2$ 1.5 mM; KCl 1,2M; Tris-HCl 20 mM). The mix was left to turn on a rotator (30 minutes, 4° C), before being centrifuged (at maximal speed, 4° C). The supernatant containing the nuclear proteins was recuperated and stored at -80° C.

11. Protein quantification

All proteins were quantified by Bradford assay. The reaction mix contained 20% Bradford reagent (BioRad) and 2% of protein extract. The assay was mixed, incubated at room temperature for 5 minutes and measured for optical density (600 nm). The result was compared to a previously established standard curve.

12. Western blots

12.1. Protein separation by gel electrophoresis

In preparation for the electrophoresis proteins were mixed with 1x loading buffer (Laemmli buffer 5x: 10% SDS; 50% glycerol, 25 % β-Mercaptoethanol; 300 mM Tris-HCl pH 6.8; 0.04% bromophenol blue). Before being loaded onto the gel, the protein buffer mix was heated (95° C, 2 minutes) to denature the protein. The gel was divided into a stacking gel (4% acrylamide/bisacrylamide 29:1; 0.125 M Tris-HCl pH 6.8; 0.1% SDS; 0.1% APS; 0.1% TEMED) and a separating gel (X%acrylamide/bisacrylamide 29:1; 0.375 mM Tris-HCl pH 8.8; 0.1% SDS; 0.1% APS; 0.1% TEMED) with variable acrylamide concentrations depending on the size of the protein to be analyzed. The gel was run at 100 V for approximately 3 hours immersed in running buffer (25 mM Tris pH 8.3; 200 mM Glycine; 0.1% SDS).

12.2. Protein transfer on to a nitrocellulose membrane

To transfer proteins from the gel onto the membrane both semi-dry and immersion transfers were performed. In both cases the gel was placed on the nitrocellulose membrane (Amersham Protran 0.45 NC, GE Healthcare) and sandwiched between two layers of three Whatmann papers each. The transfer buffer used was the same for semi-dry and immersion transfers (Tris 25 mM pH 8.3; Glycine 0.2 M; ethanol 20%). The gel stack was then either placed on a Trans-Blot Semi-Dry Transfer Cell (BioRad) (15V, 2 hours) or into a Mini-PROTEAN-Tetra Cell (BioRad) (30V, overnight, 4° C).

12.3. Blocking and antibody incubation

To test for correct protein transfer, the membrane was washed (5 minutes under agitation) with a rouge ponceau solution (1% w/v, 5% acetic acid v/v). Size markers were noted on the membrane, which was then washed three times (10 minutes under agitation) with TBST (Tris 10 mM pH7.5; NaCl 150 mM; 0.1% Tween 20). To inhibit the unspecific binding of the antibody, the membrane was washed for one hour with TBST-milk (TBST supplemented with 5% non-fat dry milk). Hybridization with the primary antibody took place over night at 4° C. The next morning the membrane was washed three times with TBST-milk (10 minutes, RT) and hybridized with the secondary antibody (1 hour, RT), which conjugated to a peroxidase. Bands were detected by incubating the membrane for 10 minutes with Clarity Western ECL (BioRad) and revealing it on film (Amersham Hyperfilm ECL, GE Healthcare).

13. MTT assay

The MTT assay measures the metabolic activity of cells. The colorless tetrazolium dye MTT (3-(4.5-dimethylthiazol-2-yl)-2.5-diphenyltetrazolium bromide) is reduced in the mitochondria to purple formazan. The absorbance measured is therefore proportional to the metabolic activity of the cells. As metabolic activity is closely linked to viability, this test is used to assess the number of living cells in a sample.

100 µL of culture medium (with or without serum) containing 6000 cells were disposed in each well of a 96 well plate. Cells were left to settle for 24h before adding 20 µL CellTiter 96 AQueous One Solution (Promega) per well to the first row of cells. After 2 hours of incubation at 37° C the absorbance was measured at 492 nm. Every day a new row of cells was thus treated with MTT and the absorbance was measured. The number of viable cells was calculated by using the formula A_x-A_1)/ A_1 with A= absorbance; 1= first day after seeding and x= current number of days after cell seeding. All measures were done in triplicate.

14. Wound healing assay

The wound healing assay is an easy inexpensive test to measure cell migration. As the name indicates, it mimics cell behavior when a wound is inflicted to a tissue. In a confluent 6 well plate a scratch is made across the cell population using a pipette tip. The time it takes the cell to close the scratch is indicative of the cells ability to migrate.

In order to prevent proliferation, cells were placed in serum free culture medium 24 hours prior to the experiment. $1x10^6$ cells / well were seeded into a 6 well plate and cultured in serum free media. Once the cells were confluent, a cross was marked in the well using a pipette tip. Immediately after a picture was taken of the cross, followed by one photo per hour, until the cross had grown over again. Wound-healing percentage of the cells was determined by the ratio of the width of the wound at each time point (t) to the wound width at t0.

15. Soft agar assay

Anchorage independent growth is one of the hallmarks of carcinogenesis. The soft agar assay is a well established method to test for a cells capacity to grow independently of a solid surface. It is considered one of the most stringent tests for tumorigenic cell growth.

To prevent the cells from adhering to the bottom of the culture dish, a base layer (DMEM 2x, 10% Fetal Bovine Serum, 100 units/mL of penicillin, 100 µg/mL of streptomycin and 2% NuSieve GTG Agarose (Lonza) was plated in the wells of a 6 well plate and left to solidify at room temperature.

3 mL of culture medium containing $1.5x10^5$ cells were mixed with 3 mL of the base layer mix and plated on top of the base. The cell agar mix was left to solidify at RT for 30 minutes and then placed in an incubator at 37° C. The following day 2 mL of culture medium were given on top of the agar to prevent dehydration.

After three weeks the cells were colored with crystal violet for 5 hours and colonies were counted by microscopy.

16. Transduction of cells with a luciferase gene

A lenti-viral vector containing a gene coding for the luciferase protein was used to transduce the MDA-MB231 and MDA-MB231-RPC32α-KO cell lines. 100 000 cells were suspended in 1 mL culture medium to which 50 µL of the lenti-viral vector (MOI of 5) were added. After inverting the tube several times the contents was given on to one well of a 6 well plate. 24 hours after transduction the culture medium was changed and the cells were amplified.

17. Test for luminescence

To verify luminescence of the transduced cells, a dilution series from 10 000 down to 39 cells per 100 µL was prepared and seeded into a 96 well plate. 18-24 hours after being plated the culture medium was replaced by 100 µL of PBS supplemented with luciferin (15 mg/mL). The cells were analyzed via the Photonimager (BioSpace Lab). Cell luminescence was plotted against the number of cells. The points measured had to fit closely to the regression line (R^2 close to 1) for the cells to be validated fur use in the *in vivo* experiments.

18. Orthotopic mouse xenografts

The cell lines transduced with the luciferase genes were xenografted into mice orthotopicaly. This made it possible to follow tumor growth and to identify metastatic sites *in vivo*.

On the day of the xenograft, MDA-MB231-luciferase and MDA-MB231-RPC32α-KO-luciferase cells were trypsinized and suspended in PBS at a concentration of 10 000 cells per 10 µL. For each cell line 10 NOD mice (NOD/Shi-scid/IL-2Rγnull) (Shultz *et al.*, 2005) were xenografted at the age of 8 weeks. Per mouse 10 000 cells were injected intraductally as described by Behbod *et al.* 2009.

Tumor progression was verified once a week using the Photonimager (BioSpace Lab). After anesthesia, mice were intraperitoneally injected with 100 µL of PBS containing 3.3 mg luciferine (E464X, Promega). Mice were monitored *via* the Photonimager (BioSpace Lab) and measurements were taken after the luminescence plateaued. Luminescence was recorded in photon per second per steradian (ph/s/sr).

After six weeks the primary tumor was removed. One month later mice were analyzed for the presence of metastasis. Afterwards mice were sacrificed and organs were placed under the Photonimager (BioSpace Lab) to measure their respective luminescence. Organs were fixated for 24 hours with 4% formaldehyde. Afterwards they were washed three times with PBS and kept in 70% ethanol.

19. Histological analysis of tumor tissue

All histological analysis were performed by the department Anatomo-cytopathologie in the Institut Bergonié.

20. Absorbance spectroscopy

All spectra were obtained using a Uvikon XL spectrophotometer. 4 µM oligonucleotide were given into in 10 mM lihtium cacodylate buffer (pH 7.2). The solution was vortexed, pipetted into quartz optical cells and overlaid with a thin layer of paraffin oil to prevent evaporation. The optical pathlength was 1 cm.

20.1. Isothermal Differential Absorbance Spectra (IDS)

Isothermal Differential Absorbance Spectra (IDS) were taken at 20° C. Wavelength was measured at 240, 260, 275, 295 and 334 nm, before and after the addition of 100 mM KCl. The spectra were obtained by calculating the difference of unfolded (without KCl) and folded (with KCl) G-quadruplexes.

20.2. UV melting analysis

Absorbance was measured between 0°-90° C at 290 nm. Constant heating and cooling rates were obtained using a Haake PG20 temperature programmer. The rate of temperature change was 12°C/h, absorbance and temperature data were taken every 6 minutes. To prevent condensation on the glass cuvette, a stream of dry air blew gently against the optical cells. Measurements were taken both during heating and cooling down, to assess for the reversibility of G4 formation.

21. Statistics

Quantitative data was analyzed using Excel. Histograms represent the average value, with error bars indicating the average deviation. Statistical analysis was performed using the programs from the website http://astatsa.com. For multiple samples an ANOVA followed by a Tukey post-hoc test was performed. Two single samples were compared using the Students t-test. Results were considered statistically significant when the p-value was $p<0,05$, with * = $p<0,05$ and **= $p<0,01$.

Chapter III

Results

The RNA polymerase III is a unique enzyme in the nuclear transcription apparatus, for it exists in two different forms. It was the discovery of RPC32α by our team (Haurie et al., 2010) that showed for the first time that two different Pol III were active in the cell, depending on the cells status. The role of RPC32α only begins to be understood. Insights to its function have been gained in stem cells (Wong et al., 2011) and transformed fibroblasts (Haurie et al., 2010). But so far no look has been taken at RPC32α in tumor cells. This work aims to bridge that gap.

1. Identification of an appropriate tumor model

The first necessity to study RPC32α, was to find a tumor model that showed a natural overexpression of POLR3G. Numerous cancer cell lines have been established for the work in the laboratory and they are freely available. The goal was to identify a cancer type that would serve as a model for the study of RPC32α. Furthermore, the hope was that RPC32α might become a diagnostic marker or even a future therapeutic target for that cancer.

To filter for cell lines that might potentially overexpress POLR3G, the databank oncogene (www.oncomine.org) was consulted. This cancer profiling database regroups transcriptomic studies from 715 datasets, with more than 86 000 samples. Breast cancer was among the cancers that showed high levels of overexpression of POLR3G. Furthermore, a study had shown that RPC32α was deregulated in a number of breast cancer cell lines used in laboratories (Neve et al., 2006).

To confirm that breast cancer would be a valid model for the study of RPC32α, we collaborated with the bioinformatic analyst Jean-Paul Feugeas (INSERM UMR 1098). Through him we were able to access transcriptomic data of 2627 clinical breast cancer samples. The different samples were grouped into 7 molecular subtypes:

- normal breast tissue
- luminal subtypes A and B
- HER2 positive
- 3 triple-negative subtypes (basal-like 1, basal-like 2 and mesenchymal)

The different subtypes were identified according to Lehman *et al.* (2011).

The data was screened for the expression levels of all Pol III subunits, as well as for some of its transcription factors and regulators. From these levels a heat map was configured (figure 20 A). POLR3G, the gene coding for RPC32α, is found to be overexpressed in all triple-negative cancers. On the other hand POLR3GL, the gene coding for RPC32β, is strongly overexpressed in the normal breast tissue, but not in the triple-negative breast cancer (figure 20 B).

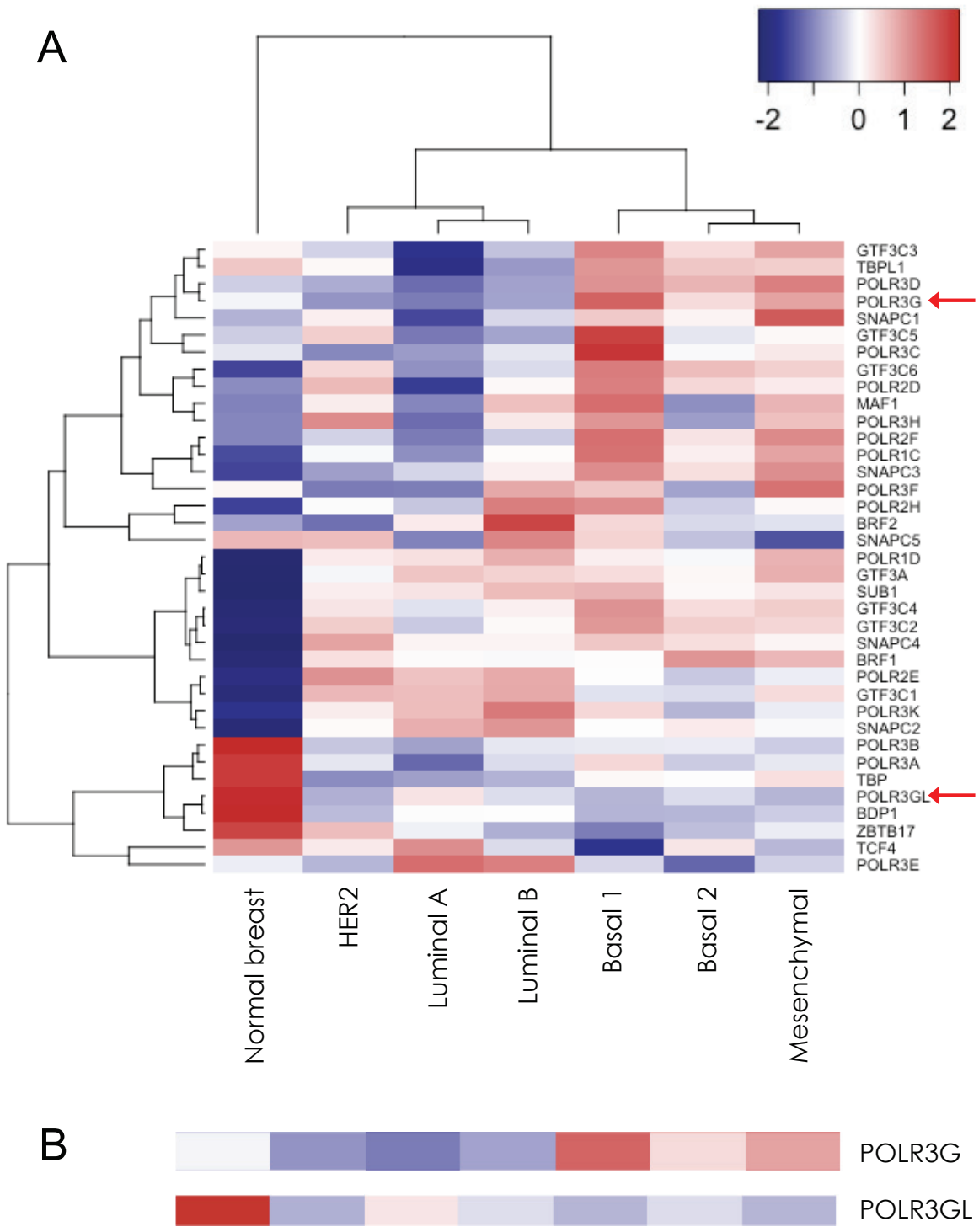

Figure 20: Expression levels of Pol III subunits and transcription factors in different molecular breast cancer subtypes. (A) Microarray data from a total of 2627 clinical breast cancer samples was gathered and analyzed. The expression levels of all Pol III subunits as well as some of its regulatory and transcription factors are shown in the heat map. (B) Enlargement of the expression levels of POLR3G and POLR3GL, genes coding for RPC32α and RPC32β respectively. While POLR3G is overexpressed (red) in the triple-negative subtypes and has low expression levels (blue) in normal breast tissue, POLR3GL shows the opposite expression pattern. Namely POLR3GL is highly expressed in normal breast tissue and has low expression levels in triple-negative breast cancer subtypes.

An analysis was performed that correlates cancer samples and the gene expression of Pol III subunits with the gene expression of Ki67 (proliferation marker), the estrogen receptor and HER2. It shows that POLR3GL is somewhat of an exception, as it is the only gene coding for a Pol III subunit that shows a clear specialization for normal breast tissue (figure 21). Besides POLR3G, there is also POLR3D, coding for RPC53, which shows some specificity for triple-negative breast cancer. However, POLR3G has a stronger bias towards triple-negative breast cancers than POLR3D. Of the other subunits that belong to Pol III, but not to Pol II, only POLR3E shows a tendency towards luminal cancers.

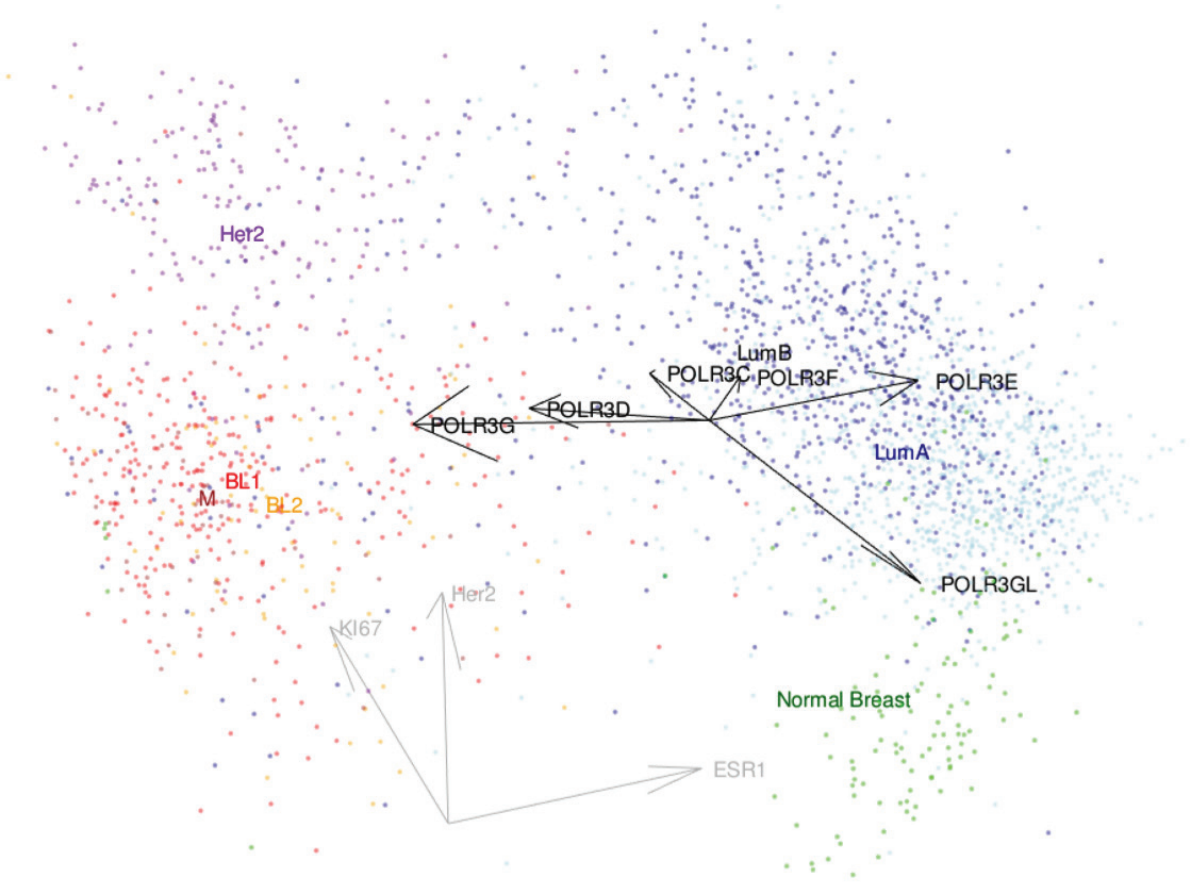

Figure 21: Correlation of Pol III subunits with different breast cancer subtypes. The genes coding for the POL III subunits that are not part of POL II were analyzed for their correlation with a certain molecular subgroup. The gene expression of POLR3G and POLR3D correlates with the triple-negative subtype, however, POLR3G correlates even more than POLR3D. The only subunit that correlates with normal breast tissue is POLR3GL. Other subtypes do not show a strong correlation for any subtype.
The proliferation marker Ki67 and the hormone receptor genes HER2 and ESR1 were used to separate the different subtypes. The dots represent the different samples analyzed; the arrows indicate the level of correlation between a gene and a certain subtype, with the length of the arrow corresponding to the level of specificity. Green, orange and red dots belong to the triple-negative subtype (BL1, BL2 and M), purple dots are HER2 positive cancers, blue dots are part of the luminal subtype (luminal A and B), and green dots indicate normal breast tissue.

The Pol III subunits shared by Pol II are found to be correlated to both luminal and HER2+ cancers, but none tends to be expressed specifically in triple-negative cancers (figure 22). The same is true for genes coding for members of the Pol III transcription machinery (figure 23 A and B). While some transcription factors are specific for HER2+ cancers, none of them shows a tendency towards triple-negative cancers.

These results show that the correlation of POLR3G with triple-negative breast cancer is not due to a general overexpression of the Pol III transcription machinery, but that it is rather a specificity of this one subunit. Therefore breast cancer and more particularly triple-negative cancers, seemed to be a suitable model to study the role of RPC32α.

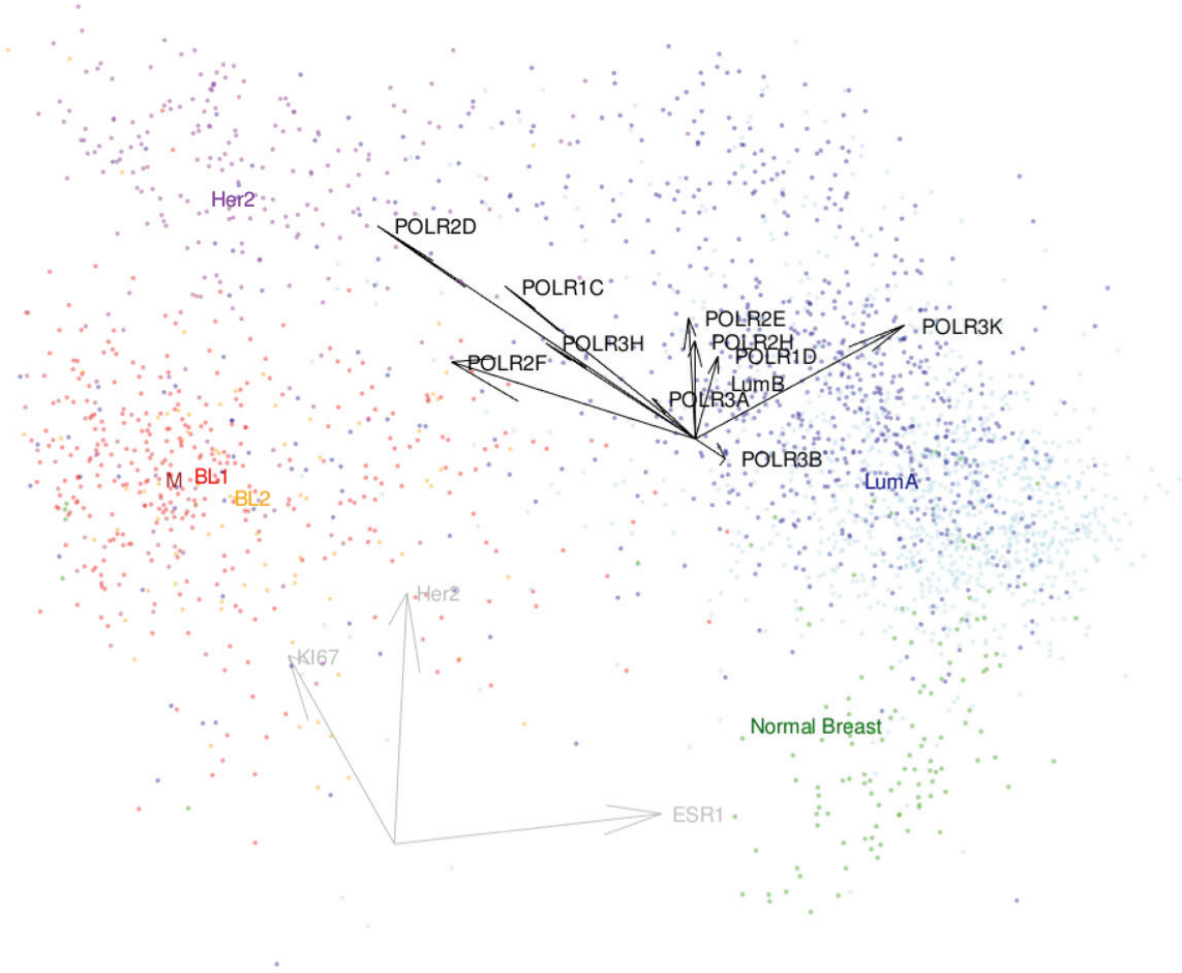

Figure 22: Correlation of Pol III subunits that have homologs in Pol II with different breast cancer subtypes. No subunit shows a strong correlation with triple-negative breast cancer or normal breast tissue. The gene expression of the different subunits correlates best with luminal and $HER2^+$ cancers.
For details on image construction see figure 21.

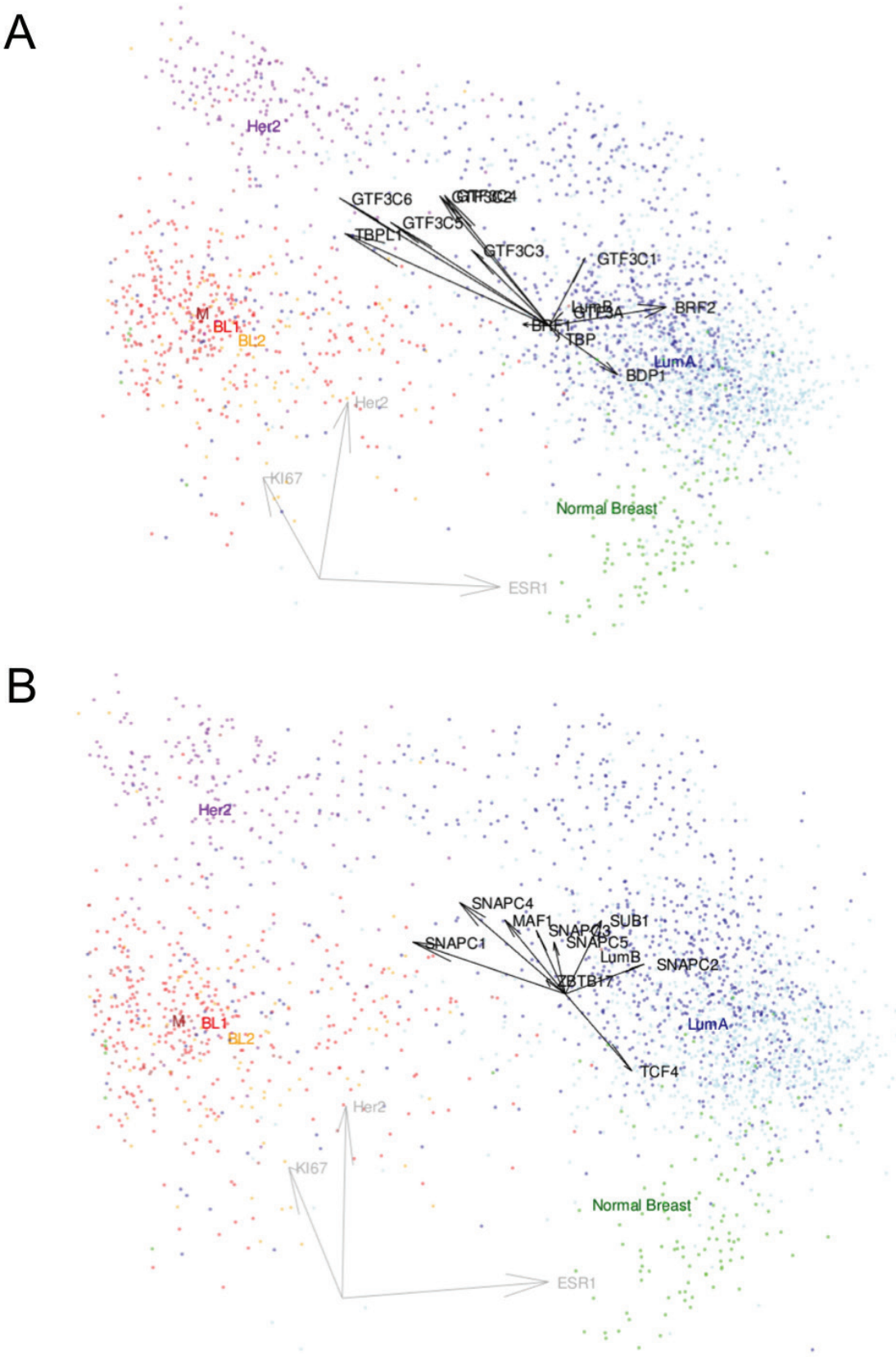

Figure 23: Correlation of Pol III transcription factors with different breast cancer subtypes. (A) Pol III transcription factors for promoter types I and II. (B) Pol III transcription factors for promoter type III. None of the transcription factors shows a close correlation with triple-negative breast cancer or normal breast tissue.
For details on image construction see figure 19.

2. Characterization of RPC32α in different breast cancer cell lines

To study the function of RPC32α in the laboratory, seven breast cancer cell lines were used (table 5). Two luminal cells lines (BT-474 and MCF7), three triple-negative cell lines (MDA-MB231, BT-549 and MDA-MB468), one HER2+/molecular apocrine cell line (MDA-MB453) and one immortalized non-tumorigenic cell line to represent normal breast tissue (MCF-10A).

Table 5: Cell lines used in this work and the breast cancer subtypes they represent

Cell line	Breast cancer subtype	Molecular subgroup
MCF-10A	normal breast tissue	
BT-747	Luminal	Luminal B (Holliday & Speirs, 2011; Neve *et al.*, 2006)
MCF7	Luminal	Luminal A (Holliday & Speirs, 2011; Neve *et al.*, 2006)
MDA-MB231	Triple-negative	Claudin-low (Holliday & Speirs, 2011) Basal B (Neve *et al.*, 2006)
BT-549	Triple-negative	Claudin-low (Holliday & Speirs, 2011)
MDA-MB468	Triple-negative	Basal (Holliday & Speirs, 2011) Basal A (Neve *et al.*, 2006)
MDA-MB-453	HER2+ / molecular apocrine	HER2 (Holliday & Speirs, 2011) Molecular apocrine (Chia, O'Brien, Brown, & Lim, 2015)

2.1. RPC32α is overexpressed in triple-negative breast cancer on an RNA and protein level

For each cell line at least three biological replicates were used to perform: RNA extraction, reverse transcription and RT-qPCR analysis. Two housekeeping genes had previously been established as a reference (RPL13A and RPL29). All data were first standardized to the two housekeeping genes and subsequently normalized to the non-tumorigenic cell line MCF-10A.

Compared to the non-tumorigenic control cell line, POLR3G is highly overexpressed in all triple-negative breast cancer cell lines (figure 24). The fold change (FC) ranges from 3.5 in MDA-MB231, over 4.5 for MDA-MB468, to 4.6 in BT-549. The other breast cancer cell lines do not overexpress POLR3G. The HER2+/molecular apocrine cell line MDA-MB453 (FC: 1.6) and the luminal cell line MCF7 (FC 1.2) have levels comparable to the non-tumorigenic control. The luminal cell line BT-474 (FC 0.8) even shows levels below that of the control. POLR3GL, the gene coding fro RPC32β, showed expression levels below those of the non-tumorigenic control.

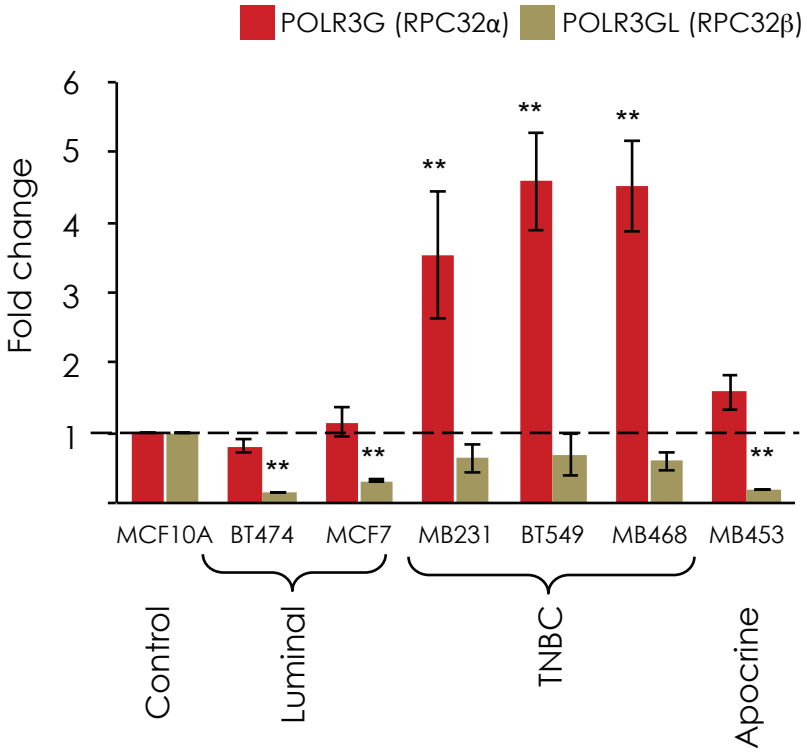

Figure 24: Expression levels of POLR3G and POLR3GL. The RT-qPCR analysis shows that POLR3G is overexpressed in the triple-negative breast cancer (TNBC) cell lines. All other breast cancer subtypes show expression levels similar to the non-tumorigenic control line MCF-10A. POLRR3GL on the other hand has its highest expression levels in the non-tumorigenic cell lines, where as all cancer cell lines show reduced expression levels. (ANOVA with Tukey post-hoc test: **=p-value<0.01).

Protein expression of RPC32α and RPC32β was measured *via* western blotting. Strong bands were detected for the three triple-negative breast cancer cell lines (figure 25). The band for MDA-MB468 seems fainter, but given that the loading control β-actin is also diminished, the expression of RPC32α is in fact similar to those in the other two triple-negative cell lines. All other cancer cell lines and the non-tumorigenic control do not show any bands. The protein levels of RPC32α corresponded to the levels of POLR3G found by RT-qPCR. Therefore it does not seem that a regulatory mechanism is active at the protein level.

RPC32β showed also bands of varying intensity. Expression is strongest in MCF-10A and MDA-MB231 cell lines and weakest in cell line BT-474. All other cell lines have a medium expression level. The only exception was the luminal cell line BT-474, which had only very faint expression levels of RPC32β. These data are similar to the expression levels on an RNA level, with two exceptions: MDA-MB231 and MDA-MB453. Both cell lines show expression levels stronger to what would be expected based on the RT-qPCR analysis. Therefore it cannot be excluded that a posttranscriptional regulation occurs or that the protein is stabilized in certain cell lines.

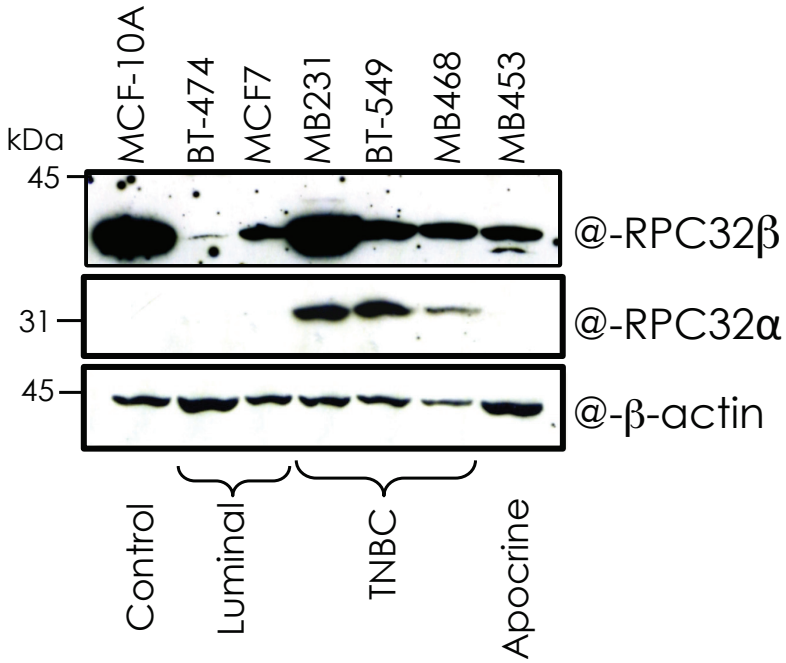

Figure 25: Expression levels of RPC32α and RPC32β. The protein levels of RPC32α and RPC32β were analyzed by western blot. RPC32α is highly overexpressed in the triple-negative breast cancer (TNBC) cell lines, while the other cancer cell lines and the control have no detectable expression. RPC32β has variable expression levels, with the highest expression levels in the non-tumorigenic control and the MDA-MB231 cell line.

Overall these results confirm the findings of the transcriptomic analysis. Namely that POLR3G is overexpressed in triple-negative breast cancer cells and this overexpression is also found on the protein level. The model of breast cancer for the study of RPC32α had therefore been validated.

2.2. No other POLIII subunits are overexpressed in triple-negative breast cancer

To find out if the overexpression of RPC32α corresponded to a general increase in Pol III activity, the RNA levels of other Pol III subunits were measured, namely of POLR3C (RPC62), POLR3F (RPC39) and POLR3D (RPC53). RPC62 and RPC39 form together with RPC32α the heterotrimer, which is unique to Pol III. RPC39 is shared by Pol I and Pol III, but does not exist in Pol II. Therefore their expression is not directly influenced by Pol II levels in the cell. The expression levels of the three subunits varied between the different cell lines. No clear pattern emerged for any single breast cancer subtype. In fact the expression levels were globally equivalent to the non-tumorigenic control, with the only exception being cell line BT-549. In the RT-qPCR analysis this cell line shows consistent high expression levels for all three subunits (figure 26 A).

However, the high RNA levels found for BT-549 are not translated into high protein levels. Western blot analysis of RPC62 did not reveal higher levels for BT-549 than for the other breast cancer lines. All of the tumorigenic cell lines show higher levels of RPC62 than the non-tumorigenic control.

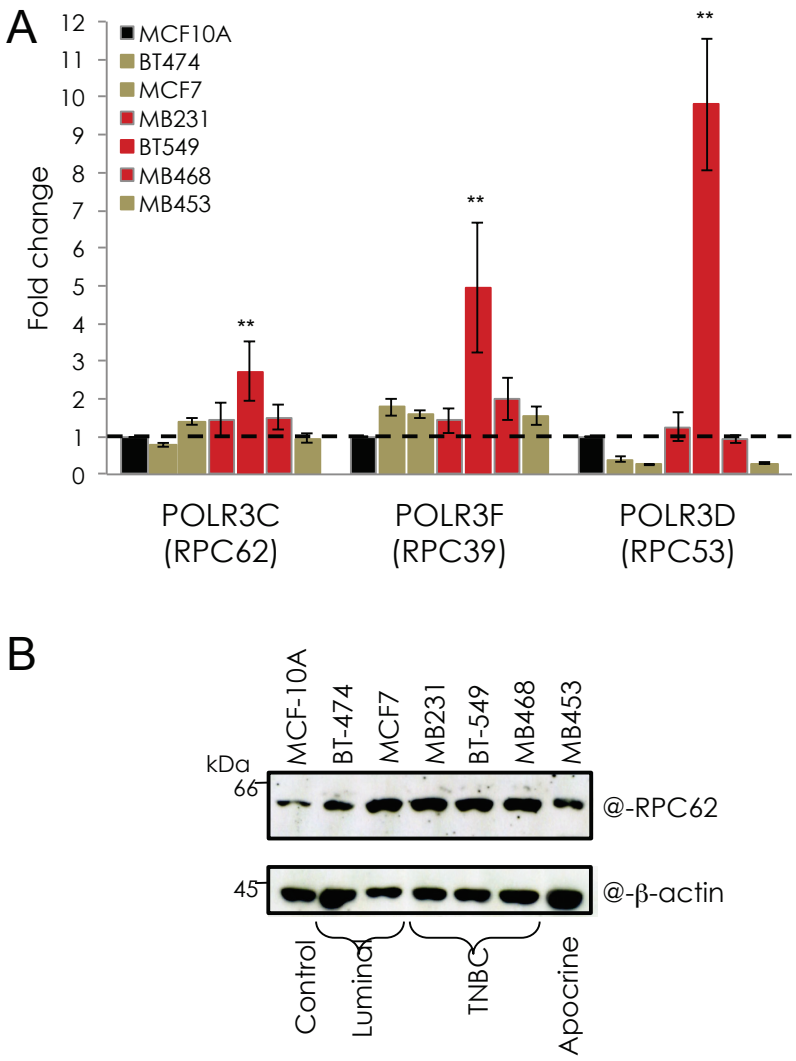

Figure 26: Expression levels of other Pol III subunits. (A) The RNA expression levels of the genes coding for the Pol III subunits RPC62, RPC39 and RPC53 were analyzed *via* RT-qPCR. The non-tumorigenic control is marked in black, the triple-negative cell lines in red, all other breast cancer cell lines are marked in beige. While the expression levels vary between the different cell lines, no clear pattern is noticeable for any one breast cancer subtype (ANOVA with Tukey post-hoc test:**= p-value<1). (B) Protein levels of the Pol III subunit RPC62 were analyzed by western blotting. All breast cancer cell lines seem to have a stronger expression than the non-tumorigenic control.

Among breast cancer cell lines neither the triple-negative nor any other subtype stands out with a significantly different expression pattern (figure 26 B).

These results indicate that while Pol III levels are elevated in all cancer subtypes, the high levels observed for RPC32α in triple-negative breast cancers are a particularity of this subtype. Therefore overexpression of RPC32α is not due to globally elevated Pol III levels. RPC32α is the only subunit that shows a breast cancer subtype specific expression pattern. Furthermore, the observations made in the different breast cancer cell lines are in concordance with the transcriptomic study on clinical breast cancer samples. This shows that the breast cancer cell lines reflect truthfully the situation in actual breast tumors. Therefore breast cancer cell lines are a valid model for the analysis of RPC32α.

2.3. RPC32α and Pol III transcripts

RPC32α is overexpressed in triple-negative breast cancer, while other Pol III subunits are not. Compared to other cancers RPC32β levels are also not particularly high in triple-negative breast cancer. Therefore the ratio of RPC32α to RPC32β and other Pol III subunits has changed. This could indicate that the high levels of RPC32α lead to an increase in Pol IIIα activity in triple-negative breast cancers. In order to find out if Pol IIIα synthesizes a particular type of Pol III transcripts, RT-qPCR was performed (figure 27). Transcripts from all three Pol III promoter types were analyzed: 5S rRNA for promoter type I, tRNAs for promoter type II, MRP RNA for promoter type III and two mixed promoters (7SL and BC200).

The expression patterns are similar for transcripts from all promoter types. This suggests that RPC32α is not specifically committed to transcription from one type of promoter. Only one cell line, MDA-MB468, shows consistent overexpression compared to the non-tumorigenic control.

Globally the triple-negative cell lines have equal or higher transcription levels than the control cell line. Only on four occasions a triple-negative cell line shows an expression level below that of the control. The cell line MDA-MB231 has a fold change lower than one for 5S rRNA (FC 0,80); MRP RNA (FC 0,77) and BC200 (FC 0,94). Once the cell line BT-549 is below the control cell line with a fold change of 0,56 for the $tRNA^{Met\ i}$. However, if the error bars are taken into account, the only instance that remains is the low levels of $tRNA^{Met\ i}$ for BT-549.

The expression levels of the luminal and HER2+ cell lines on the other hand are equal or lower than

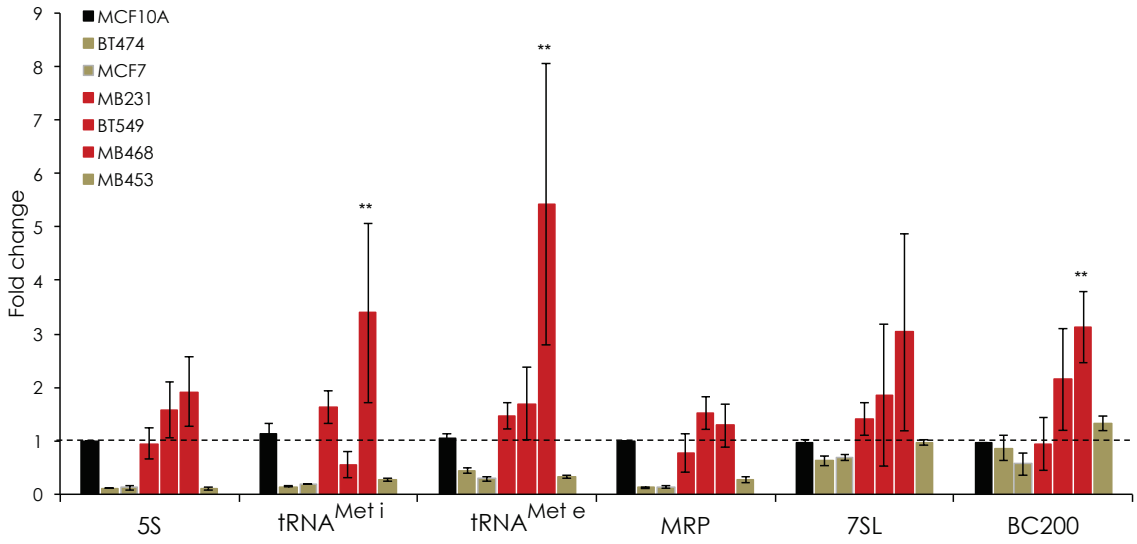

Figure 27: Expression levels of different Pol III transcripts. Compared to the non-tumorigenic control (black), triple-negative breast cancer cell lines (red) tend to have equal or higher expression levels of Pol III transcripts. The other tumorigenic cell lines (beige) generally have lower expression levels than the control. (ANOVA with Tukey post-hoc test: **= p-value<1).

the control cell line. Here the only one exception is the cell line MDA-MB453, which has a fold change of 1.32 for the BC200 RNA. The expression patterns do not correspond exactly to the amount of POLR3G measured in the respective cell lines. Therefore no specificity of RPC32α can be established. However, these results underline that triple-negative breast cancers are a group apart from other breast cancer subtypes. It remains to be elucidated if and to what extend this disparity is due to RPC32α.

3. Creation of an RPC32α knock-out cell line

RPC32α is overexpressed in triple-negative breast cancer. To analyze whether this overexpression is the cause or consequence of the cancer, knock-out cell lines were to be created. The method of choice was a CRISPR-Cas9 mediated mutation. Previous studies on RPC32α have been using shRNA to silence the expression of the protein. However, shRNA can only produce a knock-down of a protein and remaining small amounts of RPC32α might still play a role in the cell. CRISPR-Cas9 on the other hand can generate knock-out cell lines, where all functionality of the protein is abolished.

The CRISPR-Cas9 system is composed of two mandatory elements that need to be integrated into the cell: the Cas9 protein and the guide RNA with the target sequence. The target sequence is complementary to the sequence that is to be cut. It will hybridize with the DNA and thus direct the Cas9 to the target site (figure 28 A). The Cas9 will then induce a double stranded break, which can either be repaired *via* homologous recombination or non-homologous end joining (NHEJ). While the former is a safe DNA-repair mechanism, which will only very rarely produce mutations, the latter is an error prone mechanism. With NHEJ small deletions or insertions (INDELS) are a common phenomena.

For the purpose of gene editing both mechanisms can be exploited. With homologous recombination, a plasmid can be added that contains sequences homologous to the sites next to the cut. Between the homologous sequences a sequence of choice, such as a resistance to an antibiotic can be added. This sequence will be copied into the genome, when the cell uses the plasmid for homologous recombination. Cells that have been cut and repaired may then be selected using the antibiotic of choice.

However, homologous recombination in cells is a complex mechanism and the transfection with an additional plasmid might be difficult. Therefore mutations *via* NHEJ are easier to achieve, but due to the absence of a selectable marker, it becomes more difficult to identify mutated cells.

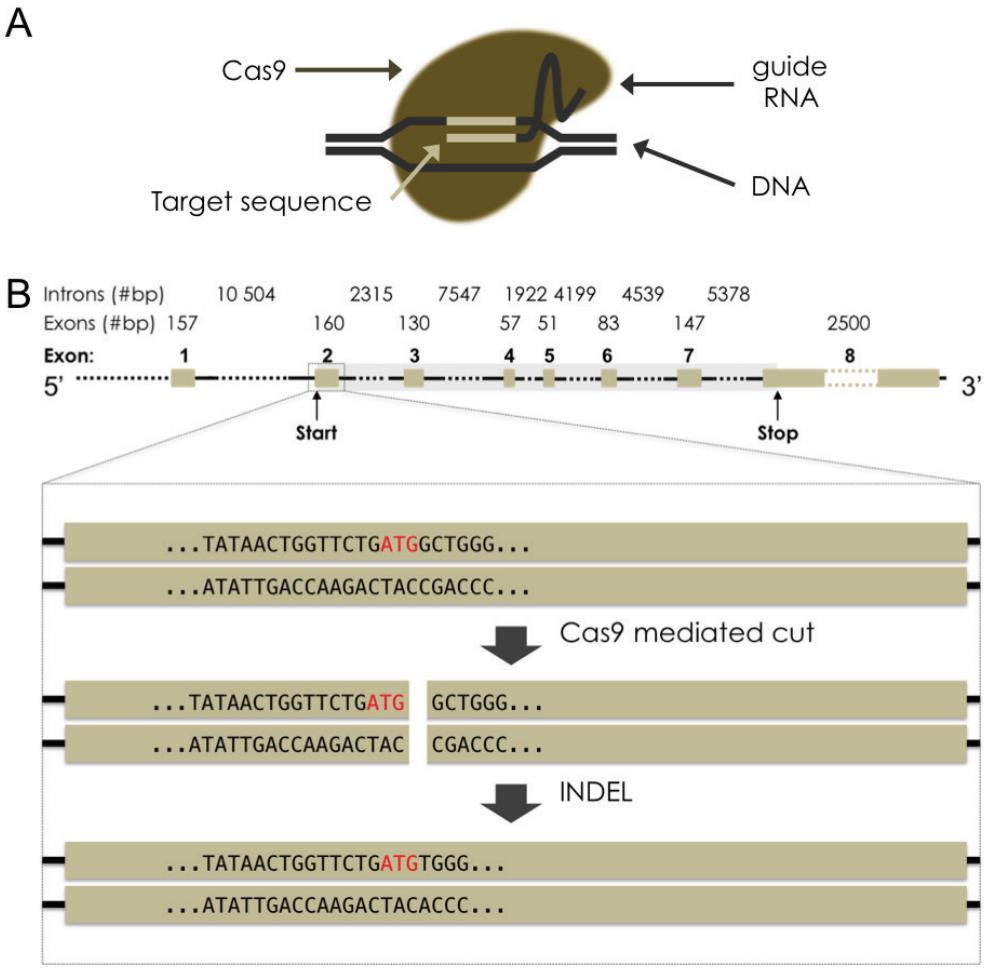

Figure 28: Mutating POLR3G using the CRISPR-Cas9 system. (A) The Cas9 enzyme is directed to the cutting site by the guide RNA. The target sequence is complementary to the sequence to be cut. The Cas9 will induce a double stranded break, which can be repaired either *via* homologous recombination or NHEJ. The latter will introduce small insertions or deletions (INDELS). (B) Schematic view of the POLR3G gene. The target sequence used for the CRISPR1 guide is located in exon 2. The cut was produced immediately downstream of the start codon (red). The double stranded break was then repaired via NHEJ, which led to small deletions.

3.1. Cutting POLR3G with CRISPR-Cas9

It was decided to create an RPC32α knock-out using MDA-MB231 as a mother cell line. In all previous analysis this cell line had shown to be an average representative of the triple-negative subtype. To obtain the knock-out a plasmid was created that contained the coding sequence for the Cas9, the guide RNA and the target sequence. The target sequence was designed so that the Cas9 would cut directly downstream of the start codon (figure 28 B). Several attempts to induce homologous recombination *via* an additional plasmid that carried a resistance to puromycine were unsuccessful. None of the transfected cells showed a resistance to puromycine.

As homologous recombination proved to be ineffective, the cells were mutated *via* NHEJ. Therefore the plasmid containing the Cas9 and the guide RNA were transiently transfected into MDA-MB231

cells. To enrich for transfected cells, puromycine was added to the culture medium for three days. Cells were then trypsinized and suspended in 1 mL culture medium. From this serial dilutions (v:v) of 1:10 000; 1:50 000; 1:100 000 and 1:500 000 were prepared. Each dilution was plated into a 96 well plate with 100 µL per well. The last dilution that still yielded colonies was used for further testing. Cells were amplified until DNA could be extracted, to test for potential frame shift mutations that would lead to a knock-out.

To test if a clone had a mutated version of POLR3G a PCR was performed, spanning the region where the Cas9 had cut. The PCR products were analyzed on acrylamide gels (figure 29 A). Candidates that showed only one band were selected to be sequenced. If the DNA sequence revealed a mutation, the sequencing profile was analyzed. Only clones that showed clear individual peaks were kept for

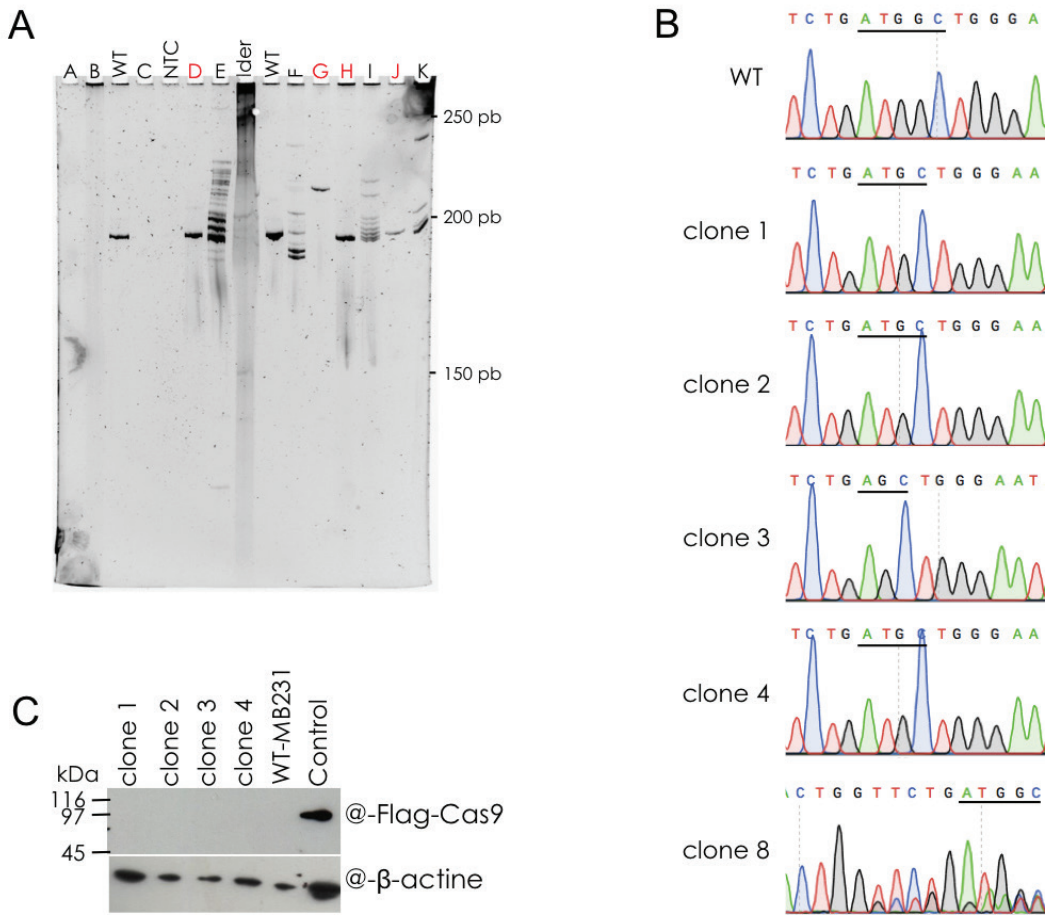

Figure 29: Creation of an RPC32α knock-out cell line. MDA-MB231 cells were transiently transfected with a plasmid containing the Cas9 and the guide-RNA. Cells were clonal diluted and the zone of interested was verified *via* PCR. (A) The PCR products were analyzed via acrylamide gels. Candidates showing multiple bands were discarded, as they are either heterozygous or the DNA came from non-homogenous cell populations. Candidates used for further analysis are marked in red.
(B) Clones that had shown only one band in the acrylamide gels were used for sequencing. The site of mutation is marked by a black line. Only clones with distinct peaks were used for further studies (clone 1-4). Clones with several peaks (clone 8) were discarded. (C) Western blots were performed in order to identify and discard clones that had stably integrated the Cas9 using an anti-Flag antibody.

further studies, as this indicates a homozygous mutation. A mix of several peaks indicates that the clone is either heterozygous or that the DNA came from a non-homogeneous cell population (figure 29 B). After several weeks of cell culture the selected clones were sequenced again, to make sure that no mixed population had been overlooked the first time.

Mutated clones that were homozygous (clones 1-4, figure 29 B), were subsequently tested for resistance to puromycine and for the presence of the Cas9. A positive answer to either test would have meant that the plasmid had been stably integrated into the cell. Only clones that were not resistant to puromycine and that had not integrated the Cas9, were used for further testing (figure 29 C). A stably integrated Cas9 would facilitate future genome editing in the cells, but there would be a risk that the Cas9 spontaneously modifies the genome.

A total of 4 clones were identified that were homozygous and that had not integrated the plasmid. Three clones show a single nucleotide deletion (clone 1, 2 and 4). They are missing a guanine downstream of the start codon (figure 29 B). Another clone has a two-nucleotide deletion (clone 3), it is missing the thymine and the guanine of the start codon. In all cases the missing nucleotides lead to a frame shift of the reading frame. For clones 1, 2 and 4 the new reading frame not only changes the amino acids, but also creates a new stop codon at position 29. In clone 3 the start codon has been mutated and therefore no translation is possible. During later analysis it was found that clone 4 had an abnormally low expression of β-actin, which is why it was excluded from further analysis.

3.2. Characterization of RPC32α knock-out cell lines

3.2.1. No feed-back loop exists in the regulation of RPC32α

The three selected POLR3G mutants were analyzed for their expression of RPC32α on an RNA and protein level. As expected the three mutant clones did not express RPC32α on a protein level (figure 30 A). Interestingly though, the RNA expression levels remain at about the same level as in the non-mutated mother cell line MDA-MB231 (figure 30 B). All three clones have POLR3G expression levels above those of the non-tumorigenic control cell line MCF-10A.

The fact that the RNA levels of the knock-out cell lines do not drop below that of the mother cell line is not a surprise. The DNA has been mutated, but the promoters and other regulatory elements are still intact. Therefore the transcription is not altered. However, it is interesting to notice that the expression levels do not rise. The lack of the protein does not seem to trigger a feed-back mechanism that increases the amount of POLR3G RNA. The cell does not seem to detect the lack of RPC32α or at least, it does not increase RNA production to counteract the loss of RPC32α.

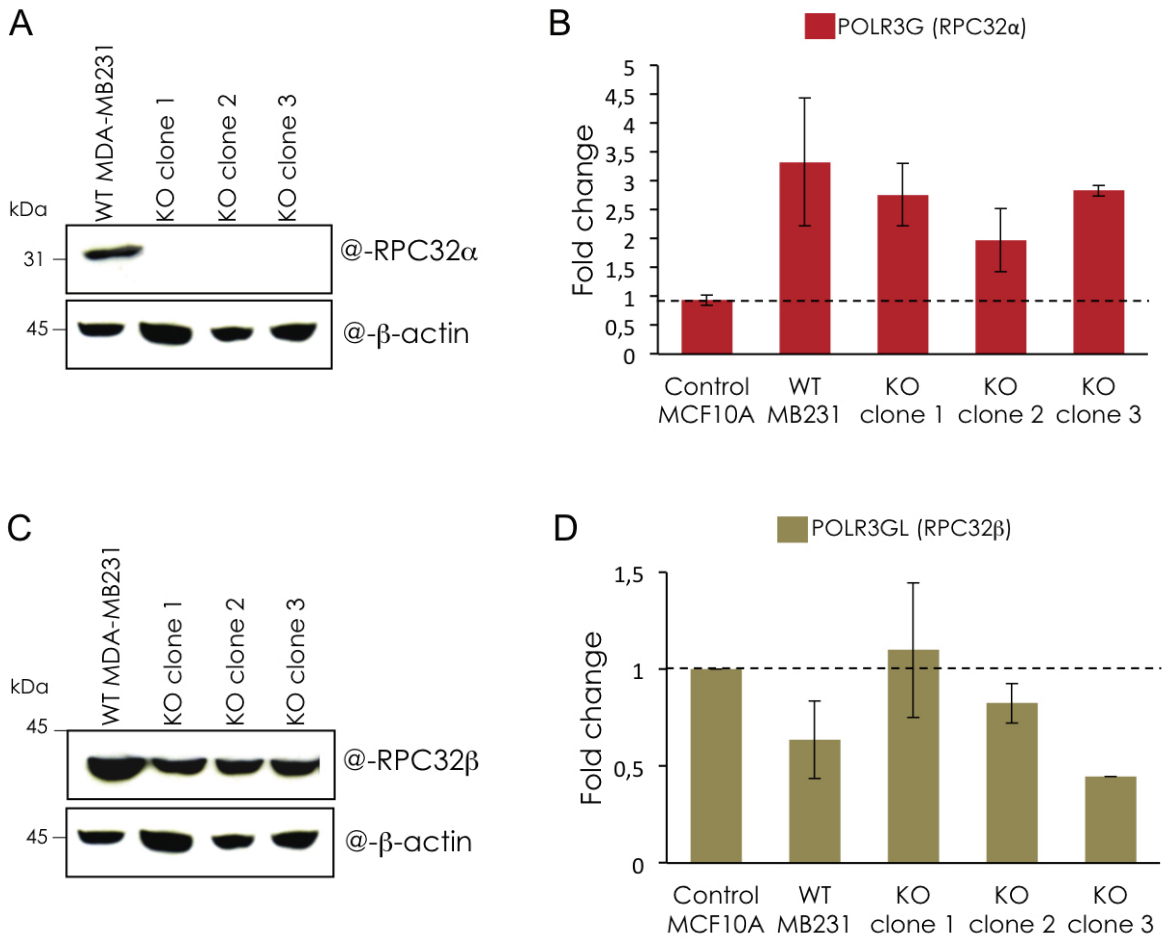

Figure 30: Expression levels of RPC32β and POLR3GL in the RPC32α knock-out cell lines. (A) Western blots of RPC32α. The three clones with a mutation in POLR3G do not express RPC32α, as expected. (B) RT-qPCR analysis of POLR3G: The RPC32α KO cell lines show similar levels of POLR3G as the MDA-MB231 mother cell line. This is expected as the start codon of POLR3G is still intact and the mutation only affects translation. (C) The western blot shows that the RPC32β KO cell lines have slightly lower levels of RPC32β than the MDA-MB231 mother cell line. Clearly the loss of RPC32α does not lead to an upregulation of RPC32β (D) On an RNA level, measured by RT-qPCR, it is observed that POLR3GL expression varies between WT and KO cell lines. However, the RPC32α KO cell lines show both over- and underexpression compared to the MDA-MB231 mother cell line. In no case the difference is significant. Therefore it does not seem that the loss of RPC32α has an effect on POLR3GL expression.

3.2.2. RPC32α and RPC32β are not co-regulated

The two homologs RPC32α and RPC32β seem to have inverse expression patterns. RPC32α is strongly expressed in tumors and stem cells, but shows very low to no expression in somatic cells. For RPC32β the opposite is true, it has high expression levels in somatic cells and low expression levels in tumor and stem cells (Haurie *et al.*, 2010; Wong *et al.*, 2011). Therefore it would have been possible that the two proteins are negatively co-regulated, when one goes up, the other goes down. However, western blots with the RPC32α knock-cell lines do not reveal an upregulation of RPC32β in the absence of RPC32βα (figure 30 C).

The same is true for POLR3GL. In the cell lines lacking RPC32α, the expression of POLR3GL fluctuates a little, but is not significantly increased compared to the mother cell line MDA-MB231 (figure 30 D). No regulatory mechanism increases the levels of POLR3GL to compensate for a loss in RPC32α. Thus it can be concluded that the expression of the two homologs is not co-regulated, neither on an RNA nor on a protein level.

3.2.3. The knock-out of RPC32α does not lead to a change in overall Pol III levels

To find out if other Pol III subunits were affected by the loss of RPC32α, the RNA levels of POLR3C, POLR3D and POLR3F were analyzed by RT-qPCR (figure 31 A). In the case of POLR3C, gene coding for RPC62, the expression levels in the three RPC32α knock-out cell lines oscillate around that of the mother cell line MDA-MB231. No significant increase or decrease can be identified. In fact it seems as if the fluctuation can be found within one cell line. But even with these variations the difference between the mother cell line and the knock-out cell lines is not significant. Furthermore, the fluctuations observed on an RNA level, are not reflected on a protein level (figure 31 B). All knock-out cell lines show expression levels similar to that of the mother cell line.

A similar picture is observed in the case of POLR3F, gene coding for RPC39 (figure 31 A). The differences observed among the knock-out cell lines seem to be due to the heterogeneity of the original cells and not to the loss of RPC32α. For no clear trend emerges among the knock-out cell lines, while some show slightly higher expression levels than the mother cell line, other show slightly lower levels. Therefore, the overall difference between knock-out and mother cell lines is not significant.

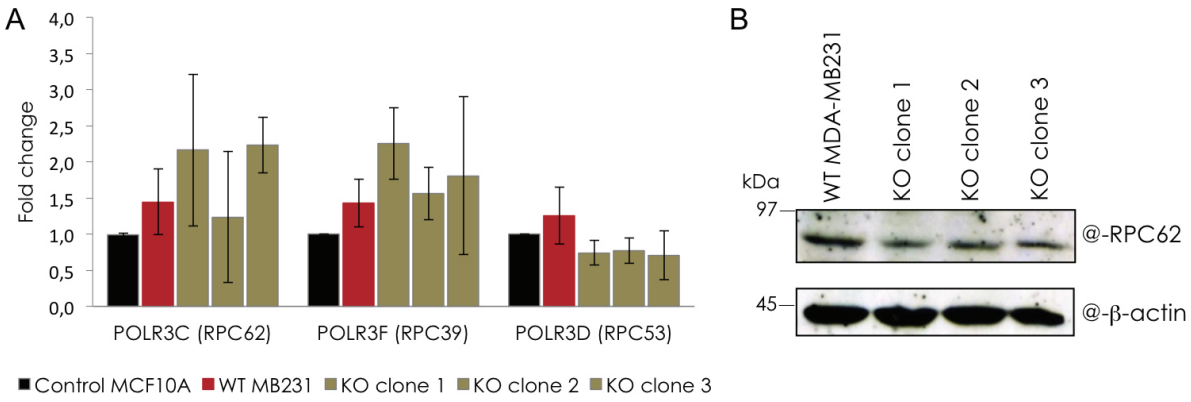

Figure 31: Expression levels of other Pol III subunits in RPC32α knock-out cells. (A) The RNA levels of subunits RPC62, RPC39 and RPC53 are not significantly altered compared to the mother cell line MDA-MB231. (B) Western blot of RPC62 in different RPC32α knock-out cell lines. The levels of RPC62 are at about the same level as that of the mother cell line.

Unlike the two other subunits, POLR3D displays a uniform pattern for all knock-out cell lines (figure 31 A). All knock-out clones show a slight decrease compared to the mother cell line. However, the decrease is small and not statistically significant. These results show that the loss of Pol III does not alter the expression levels of other Pol III subunits. This means that overall Pol III levels do not change in reaction to the absence of RPC32α.

3.2.4. RPC32α knock-out cells show reduced transcription of some, but not all Pol III transcripts

As RPC32α is part of an RNA polymerase, it seemed possible that the loss or RPC32α might alter the expression levels of Pol III transcripts. The levels of transcripts from all Pol III promoter types were analyzed *via* RT-qPCR (figure 32). The levels of almost all transcripts examined (5S rRNA, $tRNA^{Met\ i}$, $tRNA^{Met\ e}$, MRP RNA and 7SL RNA) were reduced significantly. The reduction occurs both compared to the mother cell line and to the non-tumorigenic control.

These results could lead to the conclusion that the loss of RPC32α induces a general reduction of Pol III activity. However, one transcript, BC200, does not show a significant reduction neither compared to the mother cell line nor to the control. Interestingly there seems to be a lot of volatility in the expression levels of BC200, both in the mother cell line and in the knock-out cell lines. But globally the expression levels of BC200 in the knock-out clones are not significantly different from the that of the mother cell line. This proves that it is not the overall Pol III activity that is

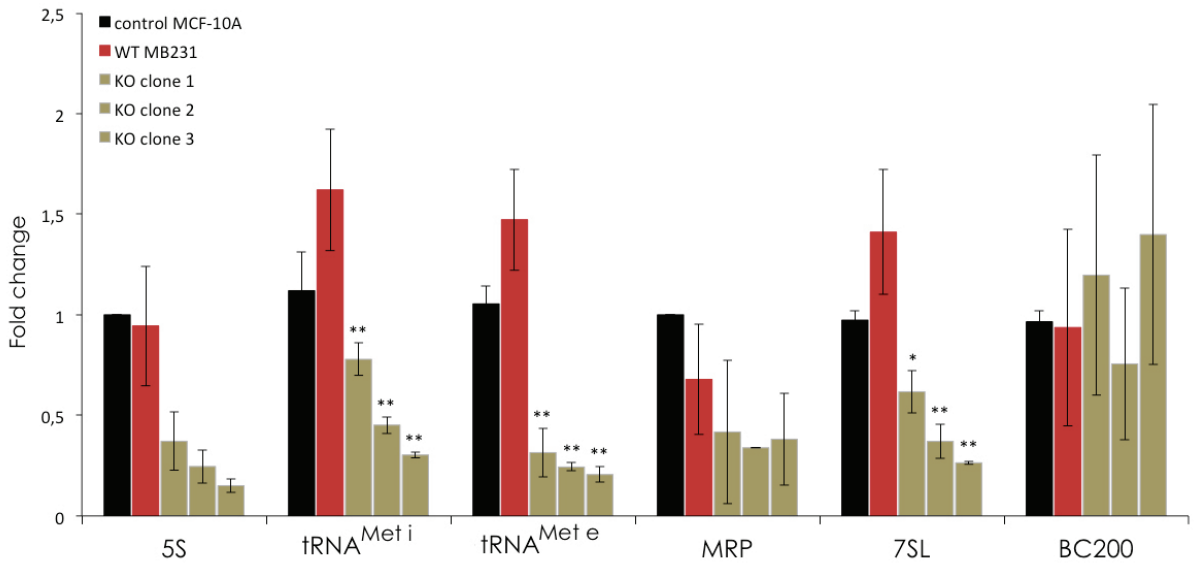

Figure 32: Expression levels of different Pol III transcripts in RPC32α knock-out cells. The RPC32α knock-out cell lines (beige) show reduced expression levels for a number of Pol III transcripts, both compared to the mother cell line (red) and the non-tumorigenic control (black). The only exception is BC200, which shows volatile expression levels both in the mother cell line and the knock-out cell lines, but does not have reduced expression levels in the knock-out cell lines. (ANOVA with Tukey post-hoc test: * = p-value<0,5; ** = p-value<0,01 compared to WT MB231.)

downregulated, but rather that of a select group of transcripts. The fact that not all transcripts are affected by a loss of RPC32α indicates that there seems to be some specificity of transcription between Pol IIIα and Pol IIIβ. Transcripts that are preferentially synthesized by Pol IIIβ naturally are less concerned by a loss of RPC32α and maybe BC200 is part of these transcripts.

3.2.5. RPC32α knock-out cell lines do not have an altered phenotype or rate of proliferation

The fact that the expression levels of several Pol III transcripts are down regulated, posed the question, whether this would lead to a change in the cells phenotype or growth rate. For among the altered transcripts are 5S rRNA and tRNAs, which are essential for growth and proliferation. However, under the microscope the knock-out cell lines present the same phenotype as the mother cell line (figure 33 A).

To test whether the cells showed altered proliferation rates, MTT assays were performed. The MTT assay uses the capacity of mitochondria to reduce the colorless tetrazolium dye MTT (3-(4.5-dimethylthiazol-2-yl)-2.5-diphenyltetrazolium bromide) to purple formazan. The metabolic activity of mitochondria can therefore be measured *via* absorbance. As mitochondrial activity is closely linked to viability, the MTT can be used to assess the number of living cells in a sample and thus over time their rate of proliferation. While there was some variability in the different clones, the overall rate of proliferation of the knock-out clones was not different from that of the mother cell line (figure 33 B).

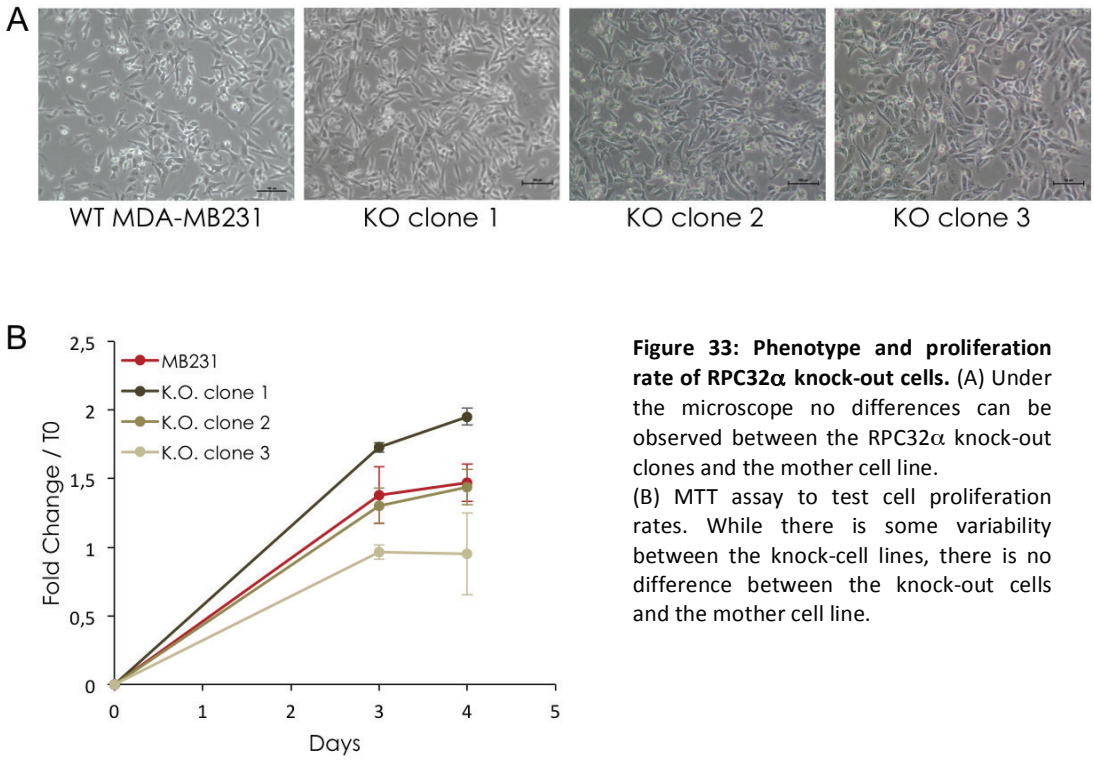

Figure 33: Phenotype and proliferation rate of RPC32α knock-out cells. (A) Under the microscope no differences can be observed between the RPC32α knock-out clones and the mother cell line.
(B) MTT assay to test cell proliferation rates. While there is some variability between the knock-cell lines, there is no difference between the knock-out cells and the mother cell line.

3.2.6. Slightly altered migration capacity in RPC32α knock-out cells

One of the greatest problems in fighting cancer, are metastasis. Some tumor cells have the ability to detach from the main tumor and to migrate to distant parts of the body. The ability to migrate is therefore an indicator of a tumor's aggressiveness. One way to test for migratory capacity is the wound-healing assay. A scratch in a cell culture plate mimics a wound, which the adjacent cells will close by migrating into the scratch zone. The faster the scratch is closed, the higher the cells capacity to migrate.

Several wound-healing assays were performed under varying conditions. It was observed that the RPC32α knock-out cell lines close the wound slightly faster than the wild type MDA-MB231 cells (figure 34 A). However, after measuring the gap size at different time points, the overall difference between all knock-out cells and the mother cell line was not statistically significant. Any differences observed where probably due to clonal variations and not to the loss of RPC32α (figure 34 B).

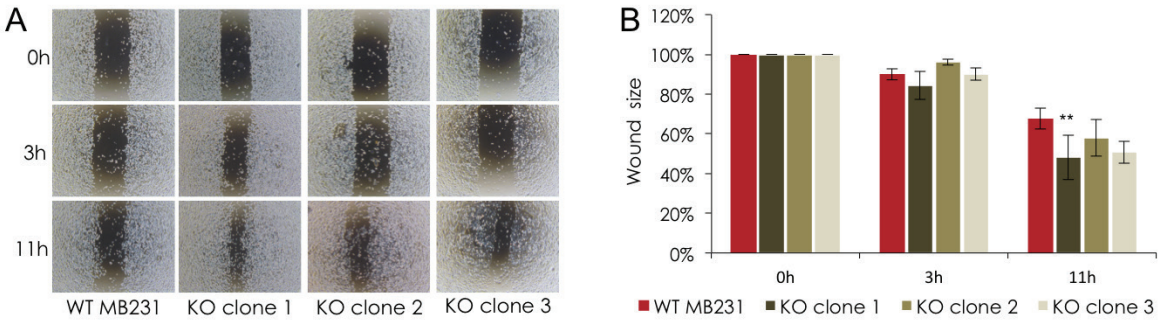

Figure 34: Wound-healing assay to assess migration capacity of RPC32α knock-out cells. (A) Microscopy images of the initial wound and the cell migration over time. (B) The wound closure over time was measured and RPC32α knock-out cells do close the wound faster than the mother cell line. However, the difference to the wild type MDA-MB231 line is only statistically significant for one out of three clones. (ANOVA with Tukey post-hoc test: ** = p-value < 0,01.)

3.2.7. RPC32α is important for tumorigenic growth in vitro

As RPC32α is highly expressed in tumor, but not in somatic cells, the question arose, if the loss of RPC32α would lead to a loss of tumorigenicity. One of the hallmark tests to measure tumorigenicity is the soft-agar assay. A petri dish is filled with a bottom layer of soft-agar, on which is poured a mix of soft-agar and cells. The cells are thus suspended in the agar, without the possibility to adhere to the plate bottom or to interact with surrounding cells. These conditions are too stressful for non-tumorigenic cells and they will die within days. Tumorigenic cells on the other hand are not

affected by either isolated growth or the inability to attach to a surface. They will grow, divide and form colonies in the soft agar.

The mother cell line and the RPC32α knock-out cell lines were grown in soft agar for three weeks. After a few days it became visible that the wild type cells had no problem multiplying in the soft agar, while the knock-out cells soon stopped growing. At the end of three weeks, the wild type cells had formed multiple colonies, while knock-out cells had had only formed a limited amount of small colonies. The tests was repeated three times and on average the knock-out cells formed 85% less colonies than the wild type cells (figure 35 A).

The RPC32α knock-out cells did not only form less colonies, but the size of the colonies was also diminished. A colony was counted as such, if it measured at least 30 µm. As smaller colonies were not counted the median size of the knock-out cell colonies lay between 55-87 µm (KO clone 1: 55 µm; KO clone 2: 79 µm, KO clone 3: 56 µm and KO clone 4: 87 µm). Half of the knock-out cell colonies ranged in size from 54-99 µm (KO clone 1: 54-65 µm; KO clone 2: 62-94 µm; KO clone 3: 55-59 µm; KO clone 4: 69-99 µm). Again the numbers are biased by the fact that colonies ≤30 µm were not taken into consideration.

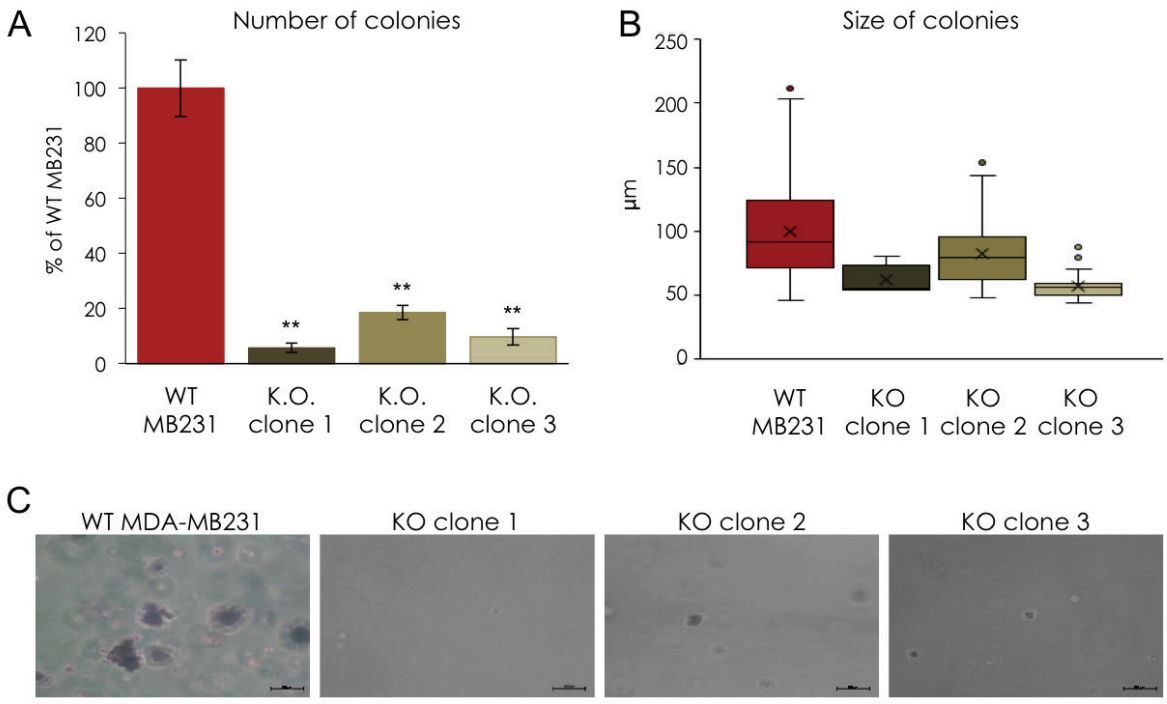

Figure 35: Soft agar assay to assess anchorage-independent growth in RPC32α knock-out cells. Anchorage-independent growth is one of the hallmarks of carcinogenesis. Cells are grown in petri dishes where they can adhere to the plate bottom. For the soft-agar assay the cells are suspended in agar and can no longer attach to a surface. Non-tumorigenic cells, cannot grow under these conditions and will die. Tumorigenic cells on the other hand will grow, divide and form colonies in the agar. (A) RPC32α knock-out cells form significantly less colonies in soft agar than the WT MDA-MB231 mother cell line. (ANOVA with Tukey post-hoc test: ** = p<0,01) (B) The size of the colonies is reduced in RPC32α knock-out cell lines. (C) Microscopy image of the different cell lines after three weeks of soft-agar growth. While the wild type MDA-MB231 cell line has formed multiple large colonies, the knock-out cell lines show only very few, small colonies.

The mother cell line had multiple colonies, which did not seem to stop growing even towards the end of the experiment. The median size of the wild type colonies was 92 µm (figure 35 B). Half of the colonies had a size of 73-124 µm. These numbers confirm what was already visible under the microscope (figure 35 C): RPC32α knock-out cells form significantly less and smaller colonies in soft-agar. Or stated otherwise: RPC32α is important for tumorigenic growth *in vitro*.

3.3. Tumorigenicity of RPC32α knock-cell lines *in vivo*

While it had been demonstrated that RPC32α is necessary for tumorigenic growth *in vitro,* it still remained unclear if the same was true *in vivo*. In order to address that question, RPC32α knock-out cells were xenografted into mice. Two possibilities exist for xenografts, either the cells are injected subcutaneously or orthotopicaly, that is at the normal site within the body. In the case of breast cancer this meant to inject the cells into the milk duct of the mice. The advantage of subcutaneous xenografts is that they are easy to inject and they are palpable even in the early tumor stages.
As tumor development depends a lot on the microenvironment, orthotopic injections can better imitate the tumor niche. It was shown that orthotopic tumors lead to increased tumor volume and better tumor vascularization (Fleming *et al.* , 2010). The downside of orthotopic injections is the complicated injection, especially in breast cancer, and the difficulty to palpate the tumor.

3.3.1. Creation of luminescent cell lines to follow tumor development in vivo

For optimal tumor development it was decided to inject the RPC32 knock-out cells orthotopicaly. In order to be able to follow tumor growth *in vivo*, the MDA-MB231 mother cell line and the RPC32α knock-out clone 1 were transduced with a viral vector, carrying a luciferase gene under the control of a murine leukemia virus-derived MND promoter (myeloproliferative sarcoma virus enhancer, negative control region deleted, *d*l587rev primer-binding site substituted).
The transduced cells were tested for efficacy and sensitivity by analyzing serial diluted cells with a photonimager. The measured luminescence given in photon per second per steradian was plotted against the number of cells. The obtained linear regression had an R^2 value of 0,9847 and 0,956 for WT MDA-MB231 and RPC32α KO clone 1 respectively (figure 36 A). This indicates that the amount of luminescence measured is proportional to the number of cells. The luminescence would therefore be a reliable indicator for tumor size. Furthermore, the luminescence was detectable for as little as 39 cells. Tumor development would thus be visible even in the earliest stages.

To test if the induction had altered protein expression or influenced the tumorigenic potential of the cells, a western blot and soft agar assay were performed (figure 36 B and C). The western blot shows that the transduced MDA-MB231-luciferase cell line still expresses RPC32α, albeit at a slightly lower level. Furthermore, the cells are still able to form colonies in the soft agar assay, at a lower number than in previous experiments. This observation is line with the lower expression levels of RPC32α.

The RPC32α-KO-luciferase cells still showed very limited ability to form colonies, at rates similar to those observed in previous assays. However, as the MDA-MB231-luciferase cells showed lower colony numbers, the difference between WT and KO cell line was diminished compared to previous experiments. These results are in line with the RPC32α expression levels observed *via* western blot. Overall the cells were validated for the *in vivo* experience.

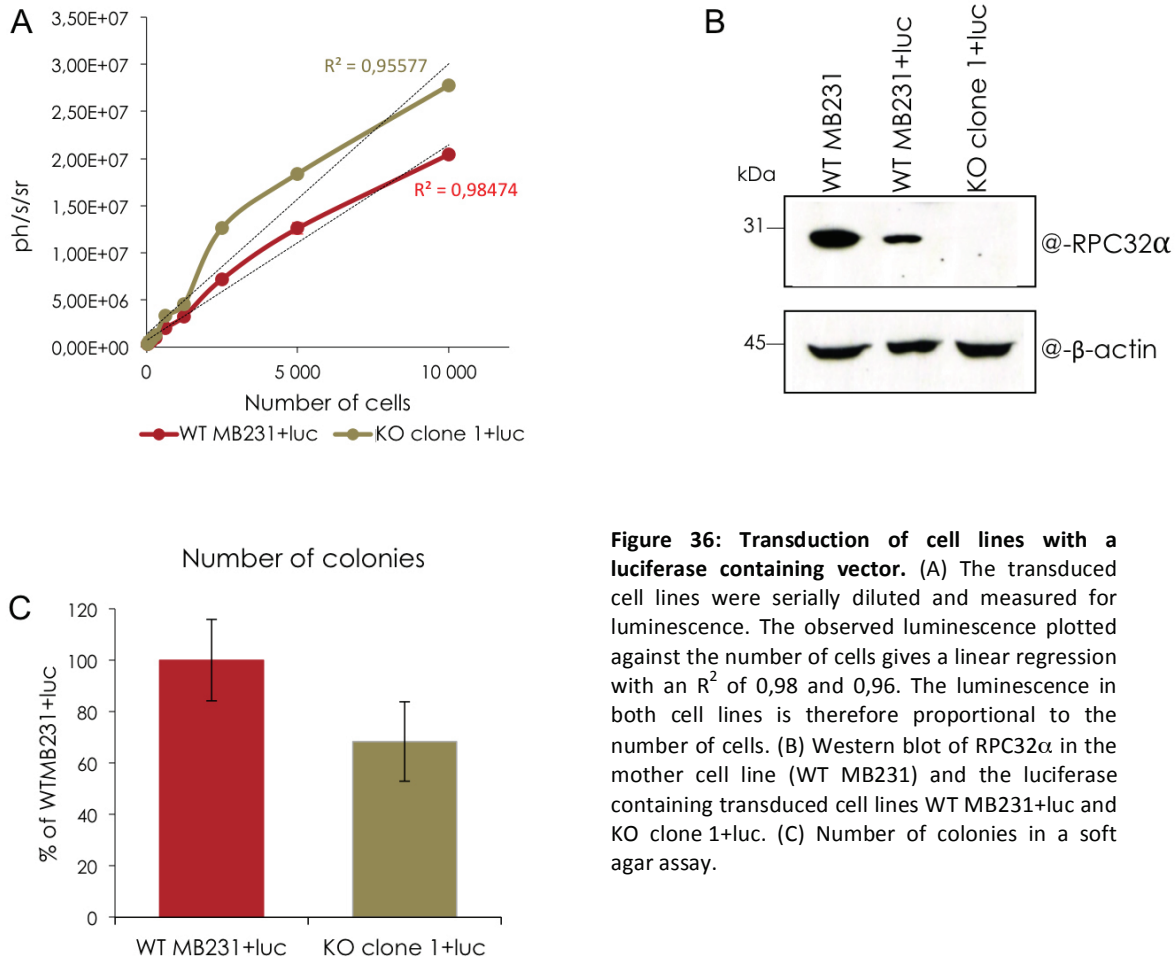

Figure 36: Transduction of cell lines with a luciferase containing vector. (A) The transduced cell lines were serially diluted and measured for luminescence. The observed luminescence plotted against the number of cells gives a linear regression with an R^2 of 0,98 and 0,96. The luminescence in both cell lines is therefore proportional to the number of cells. (B) Western blot of RPC32α in the mother cell line (WT MB231) and the luciferase containing transduced cell lines WT MB231+luc and KO clone 1+luc. (C) Number of colonies in a soft agar assay.

3.3.2. Loss of RPC32α leads to reduced tumor growth in vivo

The validated cells were injected into immunodeficient NOD mice. The injection and all subsequent analysis were performed in collaboration with Elodie Richard (INSERM U1218). Two groups of ten mice were injected with 10 000 cells per mouse. One group was injected with the transduced MDA-MB231 mother cell line, the other group with the transduced RPC32α knock-out clone 1. Tumor development was monitored *via* luminescence, measured by a photonimager. The first measure was taken one week after injection, with subsequent measures at least once a week. Tumors were left to grow for a total of 6 weeks.

All of the mice that had been injected with the MDA-MB231 cells developed tumors, which kept growing continuously until it plateaued in the last two weeks of the experiments (figure 37 A and B). The mice that had been injected with the RPC32α knock-out cells did show early luminescence, which was lost during the following weeks. For several weeks the knock-out group shows hardly any tumor growth. Only four weeks after injection does the luminescence begin to rise. During the last two weeks of the experiment the knock-out group shows steady tumor growth. The rate of growth is similar to that of the wild type group during tumor onset.

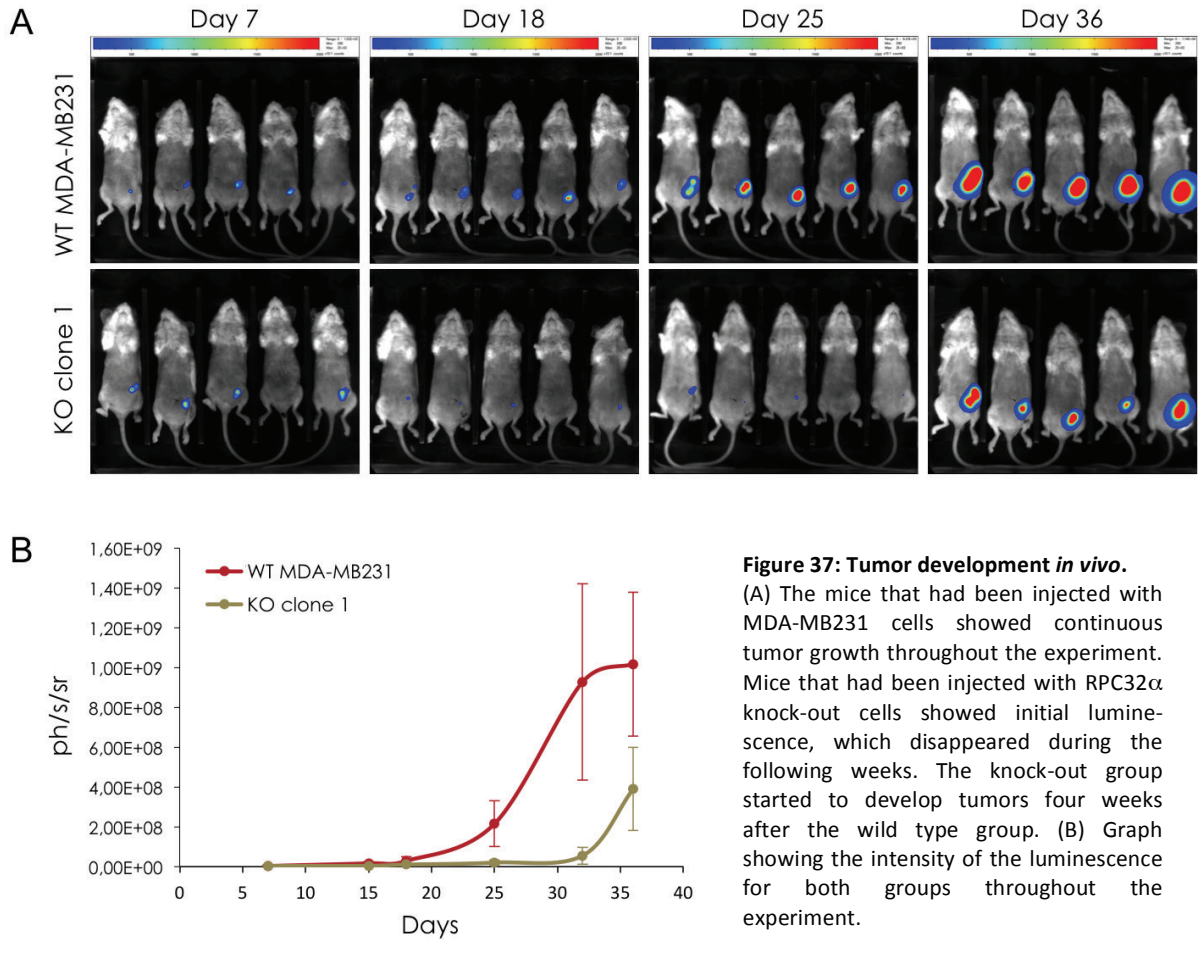

Figure 37: Tumor development *in vivo*. (A) The mice that had been injected with MDA-MB231 cells showed continuous tumor growth throughout the experiment. Mice that had been injected with RPC32α knock-out cells showed initial luminescence, which disappeared during the following weeks. The knock-out group started to develop tumors four weeks after the wild type group. (B) Graph showing the intensity of the luminescence for both groups throughout the experiment.

These result show that the loss of RPC32α leads to a highly significant (p=0,0049 at t= 15 days, p=0,0002 at t=25 days and p=0,0020 at t=36 days; students t-test) delay in tumor onset. Interestingly tumor development is not linear in the knock-out group. After initial detection luminescence decreases and remains low for several weeks. This could indicate that the injected cells were able to survive for a number of days in the mouse, but were not able to divide and grow further. When the initial cells died, the luminescence went down and only when the few surviving cells had multiplied a tumor started forming. This pattern is coherent with the observations made for the soft agar assay. There as well knock-out cells were able to survive in the beginning and form small colonies, but the cells stopped dividing in the early stages of the experiment.

After six weeks the mice experiment was stopped, at which point the tumors of the wild type group showed about double the luminescence of those in the knock-out group (p-value 0,0020; students t-test). Overall the loss RPC32α led to a significantly delayed tumor onset and significantly reduced tumor size.

3.3.3. The RPC32α knock-out cells generate significant less metastasis

RPC32α is overexpressed in stem cells and in tumor cells, hence it was of interest to see if RPC32α also was important for tumor stem cells. Among other processes, tumor stem cells play an important role in the formation of metastasis. In order to analyze if the loss or RPC32α affected metastasis formation, the primary tumor was removed and all mice were examined for metastasis four weeks later.

The mice of the wild type group had formed metastasis in multiple organs and two mice had to be euthanized before analysis to avoid animal suffering. The mice of the RPC32α knock-out group however showed only very few metastasis (figure 38 A). Overall the luminescence in the knock-out group was 100 fold less, than in the wild type group (figure 38 B).

In both groups the same organs were touched by metastasis, notably liver, lungs, pancreas and the lymph nodes. Only a few mice of the wild type group showed metastasis in the brain. No signs of metastasis in the bone were found. The number one tissue affected was the lymph nodes for both groups, followed by liver for the wild type group and lungs for the knock-out group. Overall the knock-out group showed less than 1% of the luminescence of the wild type group (figure 38 C). In other words, the wild type group had a level of luminescence 100 times higher than the knock-out group. The only exceptions were the lymph nodes, where the knock-out group had a level 5% that of the wild type group. In conclusion it can be noted that the loss of RPC32α leads to drastically reduced levels of metastasis.

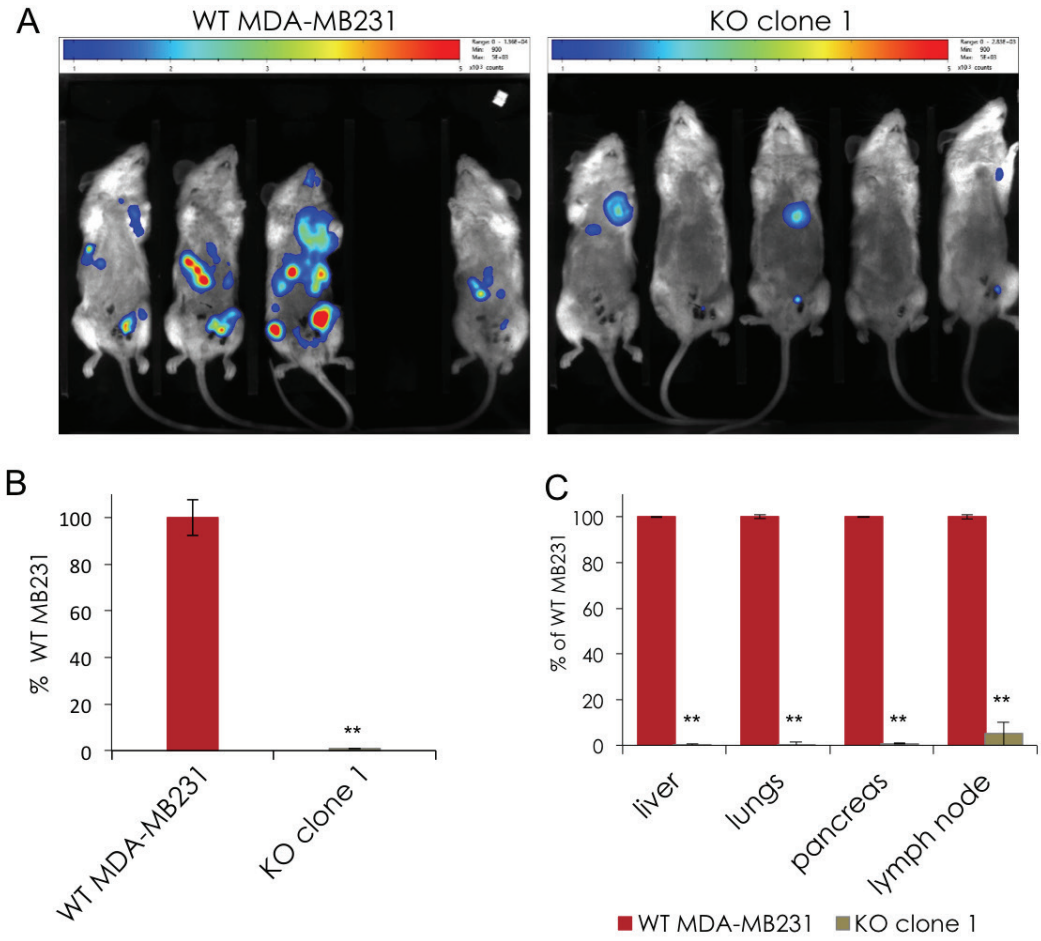

Figure 38: Metastasis in the wild type and knock-out group. (A) While the wild type group had many metastases the knock-out group showed only very few metastases. (B) The overall luminescence level for metastases was 100 fold less in the knock-out group compared to the wild type group. (C) While all tissue was more affected in the wild type group, the greatest difference was found in the liver and lungs and the least difference in the lymph nodes.

These results are very encouraging and give hope that RPC32α might become of clinical use one day. However, it is important to repeat the *in vivo* experiments with at least another RPC32α KO clone from the MDA-MB231 mother cell line as well as KO clones from other mother cell lines. This is to avoid any clone or lineage specific bias.

3.3.4. Histological analysis of the primary tumor

The primary tumor that had formed in the mammary gland after injection of either MDA-MB231 or RPC32α KO cells, was ablated 6 weeks after injection. From the tumors, slices were cut and fixed in formaldehyde. The tumor tissues were histologically prepared by the department Anatomo-cytopathologie in the Institut Bergonié.

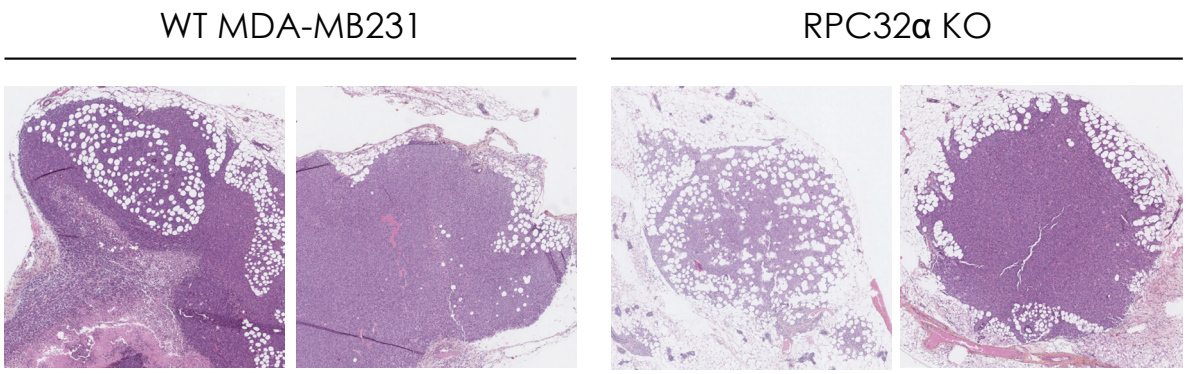

Figure 39: Histological analysis of the primary tumors. Initially the cells had been injected into the milk duct. But both tumors, those originating from MDA-MB231 cells and those from RPC32α KO cells, had started to invade the surrounding tissue.

Hematoxylin and eosin (HE) staining was used to analyze the general architecture of the tumor. Initially the cells had been injected into the milk duct, to mirror as closely as possible the natural tumor environment. HE staining showed that both tumors, those originating from MDA-MB231 or RPC32α KO cells, had left the milk duct and started to invade the surrounding tissue (figure 39).

But even if both tumors have the ability to invade the tissue around them, the extent to which they have done so differs greatly. While the mice that had been injected with the MDA-MB231 mother cell line showed tumors that had invaded almost all of the mammary gland. The mice injected with the RPC32α knock-out cells however, showed only small tumors (figure 40). This was in line with previous observations under the photonimager, which had shown that the wild type tumors were about twice the size as those formed by RPC32α KO cells.

In order to further elucidate the nature of the tumor, immunohistochemical stainings were performed. As a first step the tumor tissue was stained for human cytokeratin-7 (CK7). This marker will bind specifically to human cells and thereby permits to identify the origin of a cell in an *in vivo* mouse experiment. The staining for CK7 overlaps perfectly with the HE stained tumor cells (figure 40). Therefore the tumors observed are well of human nature. That means they originated from the cells that had been injected and are not tumors formed by the mice independent of the injection.

Given that the RPC32α KO cells had only formed small tumors, while the mother cell line was much more aggressive, the question arose, if the hormone dependence of the tumor had changed. Triple-negative breast cancers are known to be aggressive, whereas luminal cancers are more benign. Possibly the slower growth observed for RPC32α cells was due to a reacquisition of the estrogen receptor and subsequent regulation by the hormone. Consequently all tumor tissues were analyzed for the presence of estrogen receptors (ER), but both tumor groups tested negative (figure

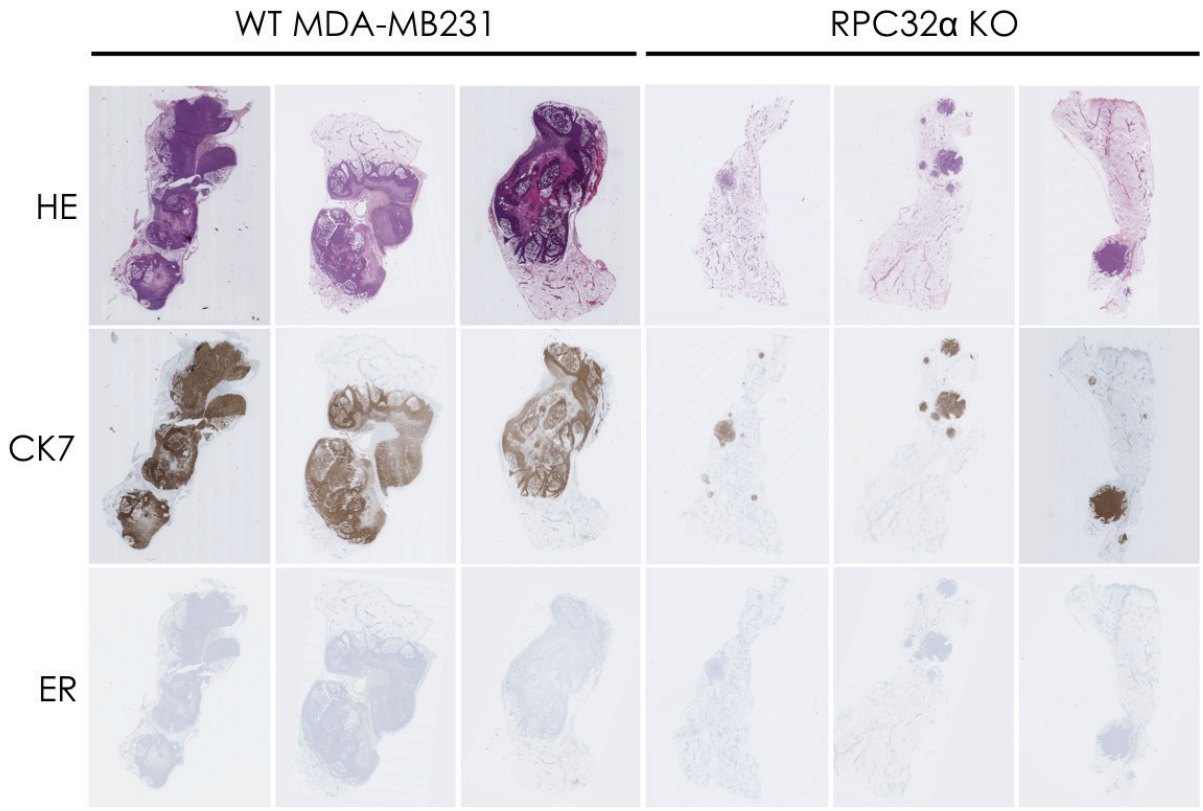

Figure 40: Histological analysis of the primary tumors. The mammary glands of the mice were removed after 6 weeks. HE staining reveals that in mice injected with the MDA-MB231 mother cell line, the tumor takes up almost all of the gland (dark violet staining). The mice injected with RPC32α KO clone show only small tumors. That the tumors are of human origin is confirmed by the CK7 staining. Both WT231 and RPC32α cells have retained their ER- status.

40). This indicates that both wild type and KO tumors had retained their triple-negative status, yet their behavior was very different.

As the MDA-MB231 wild type tumors had been more invasive and had grown to a much larger size, it seemed possible that they had a higher rate of proliferation than the RPC32α KO tumor. To test for this hypothesis the cells were stained for the proliferation marker Ki67 (figure 41). Both, the tumors formed by wild-type MDA-MB231 cells and those formed by RPC32α KO cells, showed abundant coloration for Ki67. This indicates that the proliferation rates in both tumors were similar.

The wild type tumors had caused much more metastasis, which was reason to believe that the loss of RPC32α led to a loss of stem-cellness. Cancer stem-cellness is linked to metastases, as they are the only cells among the heterogeneous tumor cells that have the ability to initiate a new tumor. A well described cancer stem cell marker is CD44. High levels of this cell surface glycoprotein, together with low levels of CD24, have been shown to be present in tumor initiating cells. Using immunohistochemistry, the wild type tumor and the RPC32α KO tumor were analyzed for the presence of CD44. Both tumors showed equivalent high levels of CD44. However, the presence of

CD44 alone is not decisive of stem-cellness. Other markers such as CD24 or ALDH1 would need to be tested to get a clearer picture.

Another mechanism that is important for the formation of metastases is the epithelial mesenchymal transition (EMT). Epithelial cells are tightly interconnected and therefore bound in place. Mesenchymal cells, on the other hand, have lost these connections and can thus detach themselves from the tumor and travel to distant sites of the body. The standard epithelial marker is E-cadherin, which is expressed at low levels in stem cells. Both tumors, wild type and KO, did not show any E-cadherin expression in immunohistochemical staining (figure 41). A typical mesenchymal marker in MDA-MB231 cells is vimentin. Staining for vimentin showed that both tumor types express similar high levels of vimentin.

These results indicate that the stem-cellness of the primary tumors is unchanged. Neither the marker for stem-cellness, nor for epithelial or mesenchymal cells seemed affected by the loss of RPC32α. Yet the growth rate of the tumor and most of all the level of metastases differs dramatically

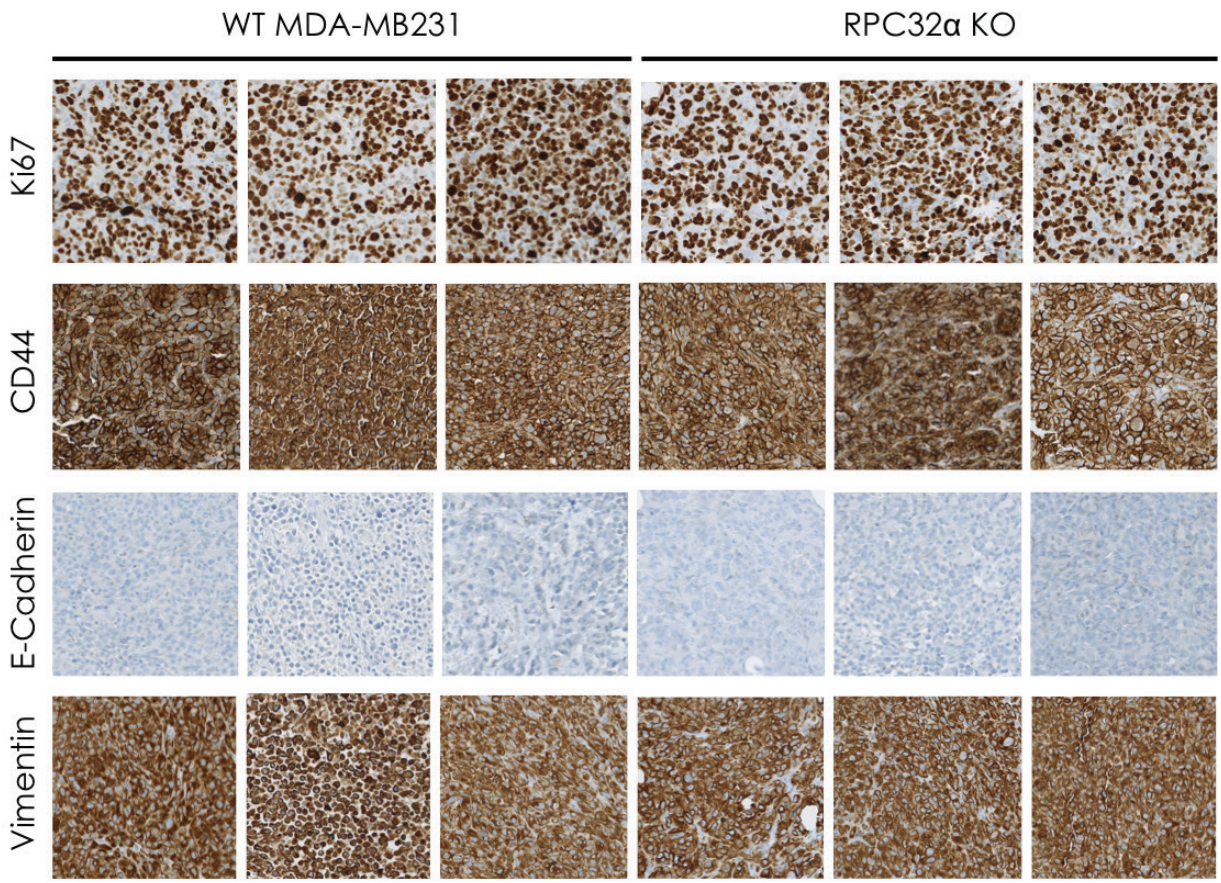

Figure 41: Immunohistochemical staining of the primary tumor. To test for proliferation the tumors formed both by the mother cell line and the RPC32α KO cell line, were stained for Ki67. No difference could be observed between the two groups. Tumor initiating capacity is linked to stem-cellness and EMT. Therefore, the tumor tissue was stained for CD44 (stem cell marker), E-cadherin (epithelial marker) and vimentin (mesenchymal marker). No difference between wild type and KO tumors could be noted for any of these markers *via* immunohistochemical staining.

between the wild type and the KO tumors. It is possible that the difference is too subtle to be detected by immunohistochemical staining. Preliminary tests using western blots did not reveal different expression levels between mother cell line and RPC32α cells. Further analysis is necessary to understand the mechanisms at work.

3.3.5. Histological analysis of the metastases

Four weeks after the primary tumor had been ablated, mice were analyzed for metastases *via* photonimager. After imaging mice were sacrificed and the different organs explored for metastases.

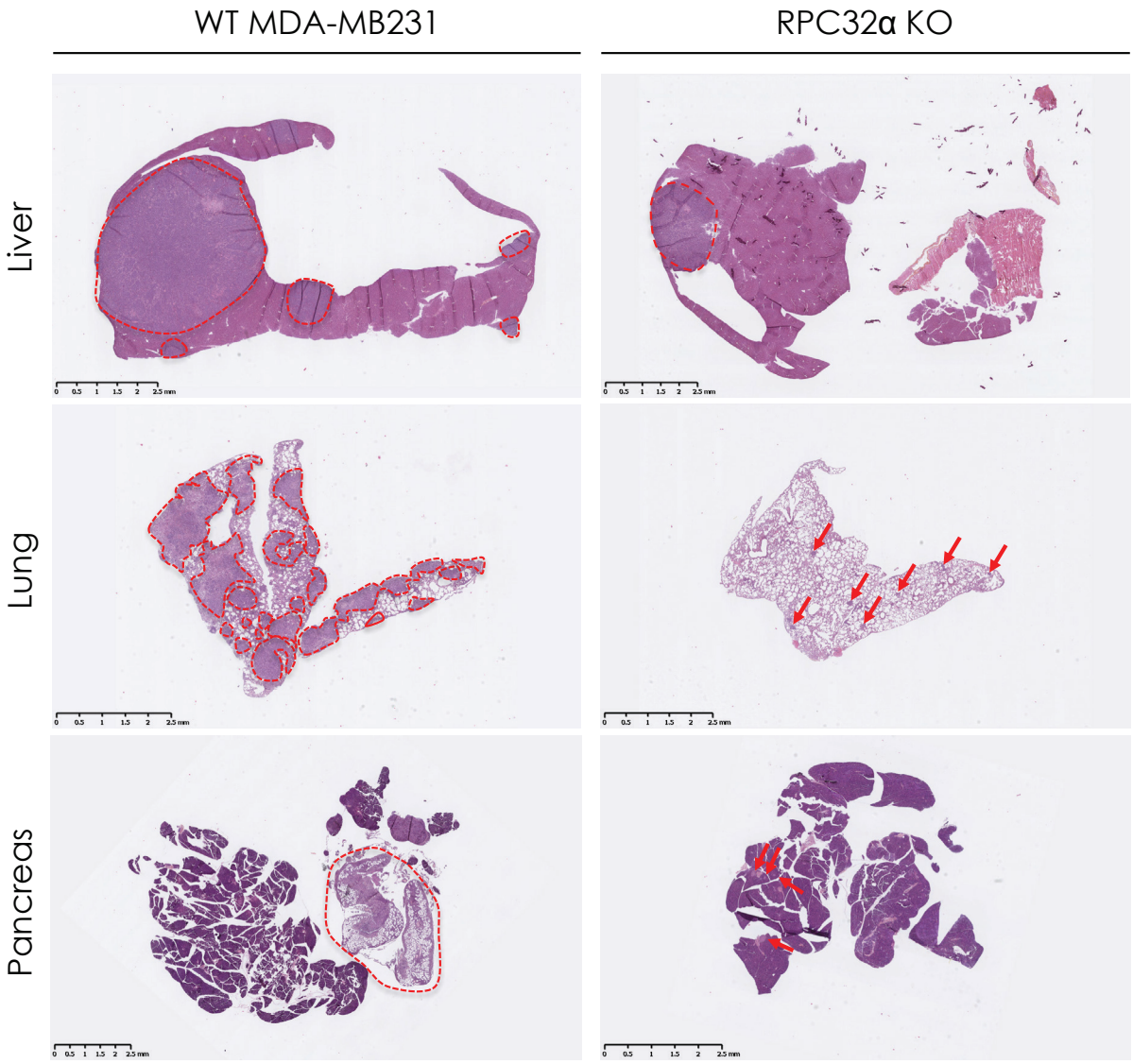

Figure 42: Hematoxylin and eosin staining of affected organs. Mice that had been injected with RPC32α KO cells showed dramatically less metastases. While the mice of the WT MDA-MB231 group have metastases in liver, lung and pancreas (red dotted lines), the RPC32α KO group had small metastases (red dotted line) and only micro metastases in the lung and pancreas. (The scale on the bottom indicates 2.5 mm.)

Slices from the affected organs were fixed in formaldehyde and sent for histological analysis to the department Anatomo-Cytopathologie in the Institut Bergonié. At the time of writing only preliminary analysis were available.

Like the primary tumor, the different organs were stained with hematoxylin and eosin (HE). In all organs analyzed the WT MDA-MB231 cell line had caused more and larger metastases. The RPC32α knock-out cells showed small metastases in the liver and only micrometastases in the lung and pancreas (figure 42). These preliminary results confirm the observations made by photonimager. Immunohistochemical analysis will be necessary to better understand the nature of the cells forming the metastases.

4. Creation of new RPC32α knock-out cell lines

The results obtained from the *in vivo* study are very promising. However, the results might be specific to the use of the cell line MDA-MB231. In order to confirm the global validity of the results, two other triple-negative breast cancer lines, BT-549 and MDA-MB468, were modified *via* CRISPR-Cas9.

One of the main critic points when using CRISPR-Cas9 are potential off-target effects. The target sequence is directed against one unique sequence in the genome and up to two mismatches are calculated in the design of the target sequence. The Cas9 is said to be very precise, but it cannot be excluded that a mismatch might occur, which would lead the Cas9 to cut at an unknown site in the genome. Therefore theoretically it would be possible that the effects observed are not the result of the RPC32α knock-out, but that they are due to an unknown mutation caused by a mismatch of the target sequence.

To exclude the possibility of a mismatch effect, the cell lines BT-549 and MDA-MB468 were mutated using the same plasmid as for MDA-MB231, but with an altered target sequence. Instead of cutting the DNA directly behind the start codon, the Cas9 was directed to cut in exon 2, 84 nucleotides downstream of the start codon. Any potential mismatch effect would be abolished, as the two target sequences would not cause the same mismatches.

Several clones were isolated for the MDA-MB468 cell line. However, in sequencing, none of the clones showed a chromatogram with discrete peaks. To exclude the possibility that the cell line is a heterogeneous cell population, cells were again serially diluted and amplified. The newly expanded subclones were tested *via* western blotting. All clones tested had reduced expression levels of RPC32α, but none of them proved to be a true knock-out clone (figure 43 A). Clone 5 shows only light traces of the protein. Using this clone for another transfection with the Cas9-plasmid, followed by limiting dilution and clonal expansion, could produce a clean knock-out cell line.

Cells from the BT-549 cell line had a lot of difficulty to grow in isolation. Therefore, after an initial selection with puromycin to enrich for cells transiently transfected with the Cas9 plasmid, cells were not directly diluted, but left to recover. After small colonies had formed on the plate theses colonies were isolated and amplified. Several clones were tested for a knock-out of RPC32α *via* western blotting. One clone was identified as an RPC32α knock-out cell line, as it showed no trace of the protein (figure 43 B). Another clone showed very light traces of the RPC32α. The two clones were sequenced and both show several peaks around the Cas9 cut site (figure 43 C). The BT-549 cell line, like MDA-MB231 and MDA-MB468, is aneuploid. The gene coding for POLR3G exists in three copies, making it difficult to have a homozygous knock-out. The fact that several peaks are visible in the chromatogram could therefore have two explanations. Either the peaks represent a heterozygous population of knock-out cells. Or clone 1 is heterozygous with all alleles mutated, but with different mutations on different alleles. Clone 2 on the other hand has at least one WT allele remaining or the cell population contains some wild type cells. In either case a novel round of CRISPR-Cas9 can eliminate any remaining wild type alleles. Nevertheless it is advisable to survey the protein expression over several cell generations, as a small number of wild type cells might remain undetected in the beginning. However, upon expansion wild type cells could amplify and make

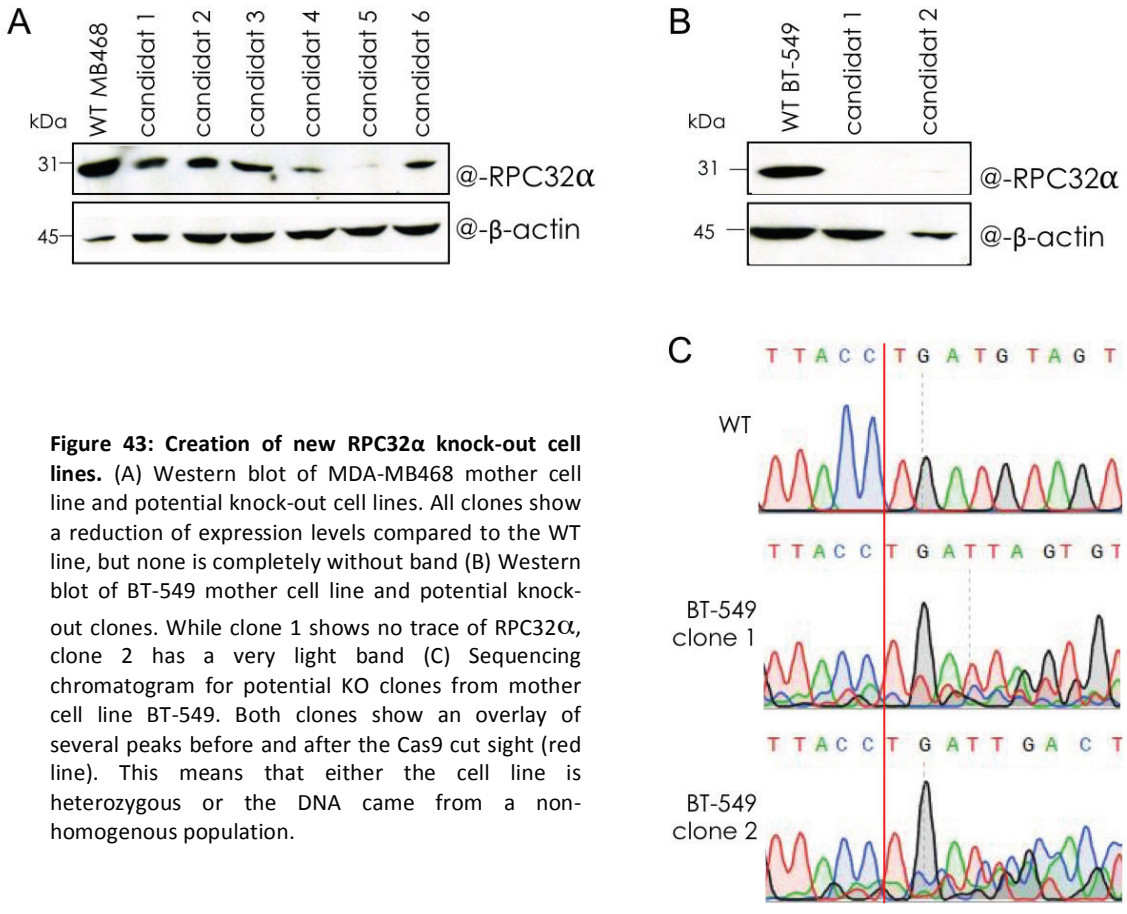

Figure 43: Creation of new RPC32α knock-out cell lines. (A) Western blot of MDA-MB468 mother cell line and potential knock-out cell lines. All clones show a reduction of expression levels compared to the WT line, but none is completely without band (B) Western blot of BT-549 mother cell line and potential knock-out clones. While clone 1 shows no trace of RPC32α, clone 2 has a very light band (C) Sequencing chromatogram for potential KO clones from mother cell line BT-549. Both clones show an overlay of several peaks before and after the Cas9 cut sight (red line). This means that either the cell line is heterozygous or the DNA came from a non-homogenous population.

RPC32α reappear. Regular testing *via* western blotting is thus recommended.

The creation of new RPC32α knock-outs, especially using different cell lines and a different target sequence, is an important step to verify the results obtained *in vitro* and *in vivo*. One knock-out clone has already been identified in the BT-549 cell line, which can be used for further experiments. Another BT-549 clone needs to be re-cut with the Cas9, to eliminate any remaining wild type alleles. No clone has been identified so far for the MDA-MB468, but some of the knock-down clones are promising candidates for a second round of CRISPR-Cas9.

5. Regulation of RPC32α

RPC32α and RPC32β are homolog proteins, yet their expression profiles are very different. Therefore regulatory elements must exist that control the expression of RPC32α, but not RPC32β. A previous study had identified MYC as a potential regulator for RPC32α, but not RPC32β (Renaud et al., 2014). The same study found that all other genes coding for Pol III subunits had a MYC binding site in their promoter. Therefore MYC dependent regulation cannot explain the overexpression of RPC32α in triple-negative breast cancer cells, as it is independent of the expression levels of Pol III. Another study reported that in stem cells POLR3G is a downstream target of NANOG and OCT4 (Wong et al., 2011). The study does not report on the effect of NANOG and OCT on POLR3GL. However, given the fact that POLR3GL is only expressed at low levels in stem cells, it seems likely that NANOG and OCT4 either do not regulate POLR3GL or regulate it in a different manner than POLR3G.

To identify potential regulators of POR3G, the promoter region and the gen sequence were analyzed. The search revealed a potential G-quadruplex (G4) structure on the non-template strand. The G4 is located in the 5' end of the first intron of POLR3G (figure 44 B) and does not exist in POLR3GL. G-quadruplexes are secondary structures that can form in guanine rich sequences. The consensus sequence for G-quadruplexes is $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$, four blocks of at least three guanines, separated by a variable number of other nucleobases. The guanines can bind *via* Hoogsteen hydrogen bonding and form so called guanine tetrads. The tetrads are often stabilized by a cation, like potassium. Several tetrads can stack up to form a G-quadruplex. A cation, most often potassium, stabilizes the structure (figure 44 A). The G-quadruplex disrupts the DNA double helix and could potentially hinder transcription and thus serve as a regulatory element.

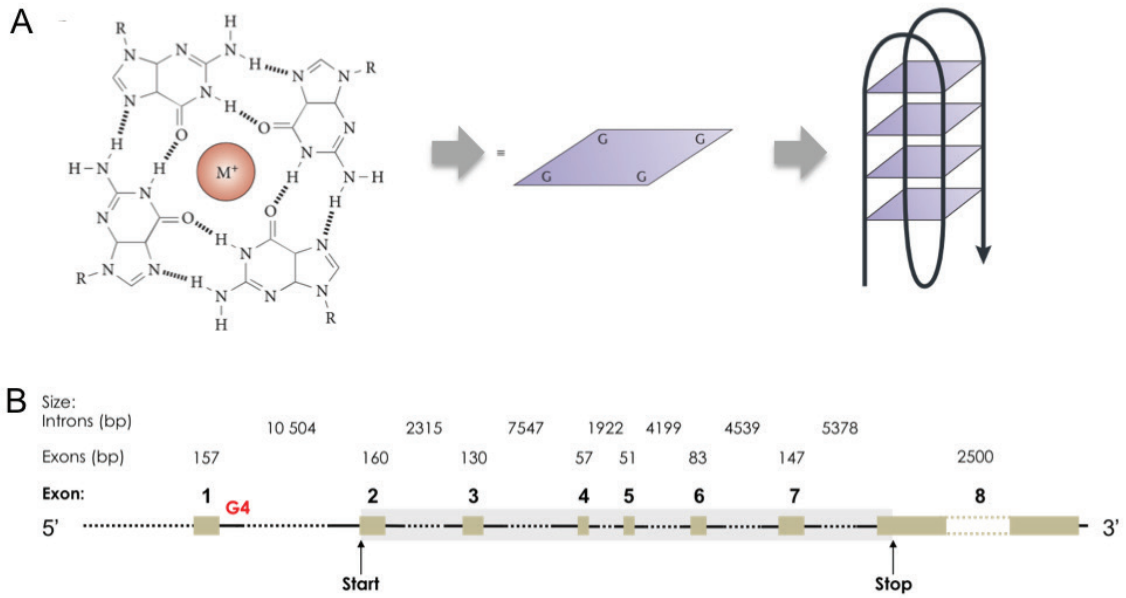

Figure 44: Potential G-quadruplex structure in POLR3G. (A) A G-quadruplex (G4) is a secondary structure in the DNA. It can form when four guanines interact to form a tetrad. The tetrads can be stack to form a G4. (B) Schematic view of the gene POLR3G indicating the position of the potential G4. The size of the introns and exons are given in number of base pairs (bp) above the gene.

The sequence identified in the first intron had five blocks of at least three guanines, plus additional guanines between the different blocks (figure 44 C). But not every sequence that is rich in guanines forms G-quadruplexes. To analyze the potential of the G-rich sequence found in POLR3G, the G4-Hunter algorithm (Bedrat *et al.*, 2016) was applied to POLR3G. Indeed the G-rich sequence in the first intron was evaluated by G-Hunter with a score of >2, which indicates a strong possibility for a G-quadruplex.

Twelve sequences were tested for their ability to form G-quadruplexes *in vitro*: the original sequence including 5 blocks of G, two sequences with either the first or last four blocks of guanine deleted as well as nine sequences with a varying number of guanines replaced by either thymines or adenines (figure 45 A). The sequences were tested using Isothermal Differential Absorbance Spectra (IDS). In this technique the absorbance of the oligos is measured at 240, 260, 275, 295 and 334 nm wavelength, before and after addition of potassium. The difference between the two spectra is shown as the isothermal differential absorbance spectrum (IDS). If a G4 has formed upon the addition of potassium, the spectrum will have shifted and several distinctive peaks are visible in the IDS. The presence of G4 is indicated by two positive peaks at 243 nm and 273 nm, as well as one negative peak at 295 nm (Mergny *et al.*, 2005). The full length sequences and the two sequences missing either the first or last block of guanines all showed typical G4 IDS peaks (figure 46). Five out of the altered sequences (numbers 4-8) were still able to form G-quadruplexes. However, sequences

A Sequence of the potential G4 identified *in silico*

TGGGAGGGGTGGTGAGGGTAGAGGGGAATTAAAGGGG

1 2 3 4 5

B Sequences used for *in vitro* analysis

Number	Sequence	G4
Sequence 1:	TGGGAGGGGTGGTGAGGGTAGAGGGGAATTAAAGGGG	Yes
Sequence 2:	TGGGAGGGGTGGTGAGGGTAGAGGGGA	Yes
Sequence 3:	AGGGGTGGTGAGGGTAGAGGGGAATTAAAGGGG	Yes
Sequence 4:	TGGGAGGGGTGGTGAGGGTAGAGGTTAATTAAAGGGG	Yes
Sequence 5:	TGGGAGGGGTGGTGAGGGTAGAGTTGAATTAAAGGGG	Yes
Sequence 6:	TGGGAGGGGTGGTGAGGGTAGATTTAAATTAAAGGGG	Yes
Sequence 7:	TGGGATGGGTGGTGATTGTAGATTTAAATTAAAGGGG	Yes
Sequence 8:	TGGGAGGGGTGGTGAGGGTAGATTTAAATTAAATTTG	Yes
Sequence 9:	TAGGATGTTTAATGAGGGTAGATTTAAATTAAAGGGG	No
Sequence 10:	TGGGATTTATGGTGAAAATAGATTTAAATTAAAGGGG	No
Sequence 11:	TAGGATGTTTATTGAAAATAGATTTAAATTAAATTTG	No
Sequence 12:	TTTTATTTATATTGAAAATAGATTTAAATTAAATTTG	No

Figure 45: G-quadruplex sequences tested. (A) The whole length sequence is composed of 5 guanine blocks with additional guanines interspersed between the blocks. (B) Twelve sequences were tested for their ability to form a G-quadruplex *in vitro*. The whole length sequence, two sequences with either the first or last four blocks of guanine deleted and nine sequences with a varying number of guanines (red) replaced by either thymines or adenines (green).

where more than one guanine block had been eliminated (number 9-12) were no longer able to form G4s (figure 45 B).

In a second step the thermal resistance of the G4s was analyzed by UV melting curves. The sequences were heated in the presence of potassium and absorbance was measured at 290 nm. Measurements started at 90° C, temperature was then decreased at a rate of 12°C/h, while absorbance and temperature were taken every 6 minutes. Arrived at 0°C the temperature increased again up to 90° C, to test for reversibility. The twelve sequences tested showed varying stability. All sequences that were able to form a G-quadruplex in the previous test, showed a high level of stability. The whole length sequence remained stable at temperatures above 40° C (figure 46). The most instable of them was number 7, which started to denature at temperatures around 25° C (figure 47). The sequences that were not able to form G-quadruplexes in the previous test consequently showed aberrant UV spectra in this test. In conclusion these tests have shown that the sequence detected in the first intron of POLR3G is capable of forming a G-quadruplex *in vitro* that would be stable at temperatures inside the human body.

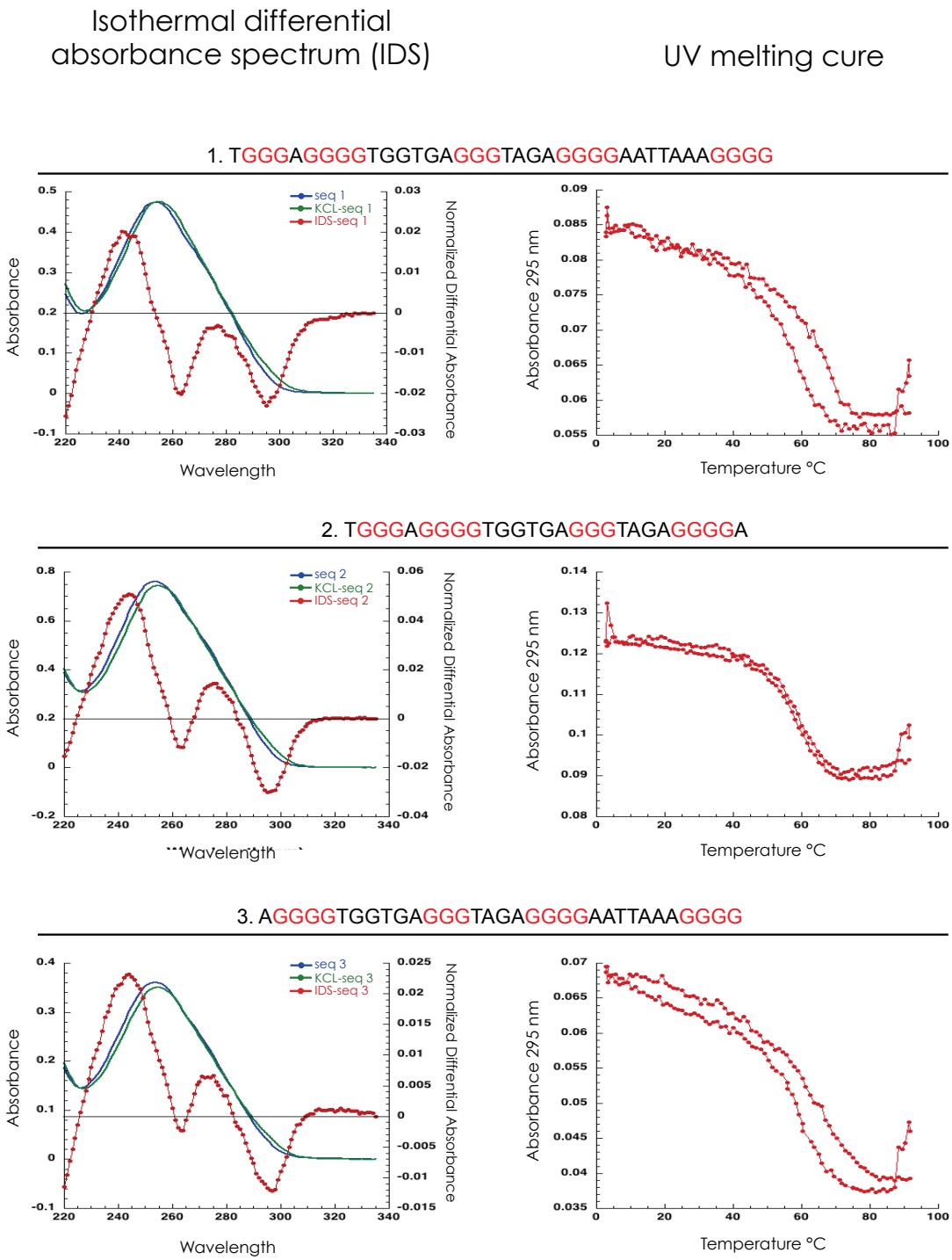

Figure 46: *In vitro* analysis of different G4 sequences. Potential G-quadruplex sequences were analyzed by isothermal differential absorbance spectra (IDS) and UV melting analysis. The whole sequence (1) and the two sequences were either the last (2) or first (3) group of guanines had been deleted show the typical G4 IDS curve, with two positive peaks at 243 and 273 nm and one negative peak at 290 nm. Proving that these sequences can form G4s. Melting analysis revealed that all G4s were stable up to temperatures above 40° C. This means they could form in the human body.

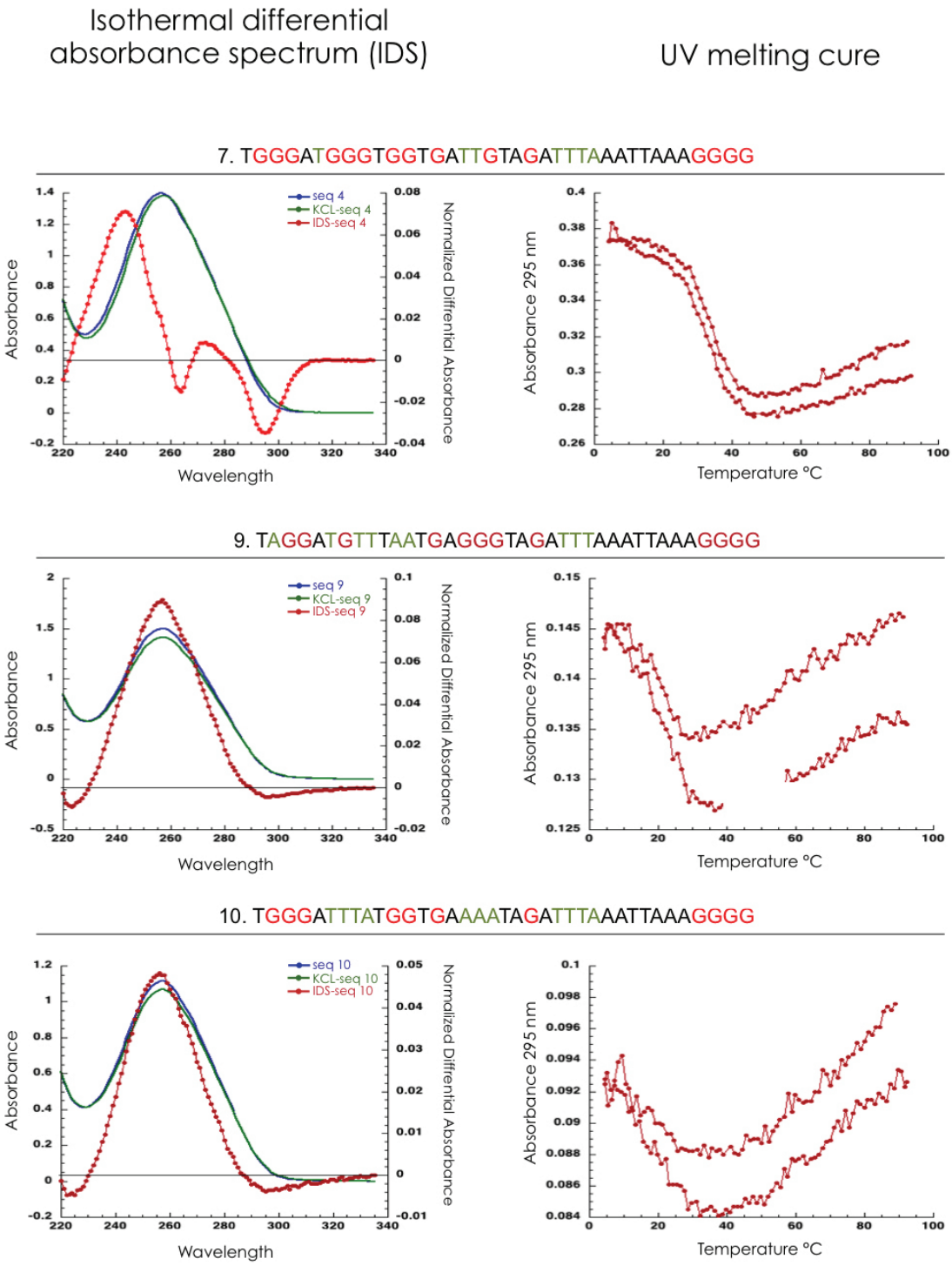

Figure 47: *In vitro* analysis of mutated G4 sequences. The original G4 sequence had been mutated by replacing guanines (red) with either thymines or adenines (green). Sequences with only a few mutations were still be able to form aG4 (sequence 7), but sequences which had more than one guanine block deleted (sequences 9 and 10) do no longer show the G4 IDS profile. Thermal melting analysis reveals that while sequences 7 can still form a G4, it is less stable than the whole length sequence. Sequences 9 and 10 cannot form a G4 and therefore show aberrant melting curves.

Once it had been established that the G-rich sequence in the first intron could indeed form a G-quadruplex and that it would be stable at around 37°C, the question arose of the possible function such a structure might have. One hypothesis is that the G4 could block transcription of the gene, thereby regulating its expression. Pol II would start transcription on exon 1, but abort as soon as it reached the G4. If this was the case than small parts of RNA that correspond to the first exon should be in access compared to the other exons. Therefore RT-qPCR analysis was performed using primers in exons 1, 6 and 8 (figure 48 A).

As the overall expression levels of POLR3G in the different cell lines vary a lot, it was necessary to normalize the expression levels of the exons. Both exon 1 and 8 were normalized by dividing their expression levels by that of exon 6. This made it possible to immediately compare the expression between the different cell lines. In the three triple-negative breast cancer cell lines (MDA-MB231, BT-549 and MDA-MB468) show elevated levels of exon 1, compared to their exon 8 expression. In all other cell lines the expression of both exons is approximately the same (figure 48 B).

The three cell lines that have elevated exon 1 levels are also the three cell lines that have the highest

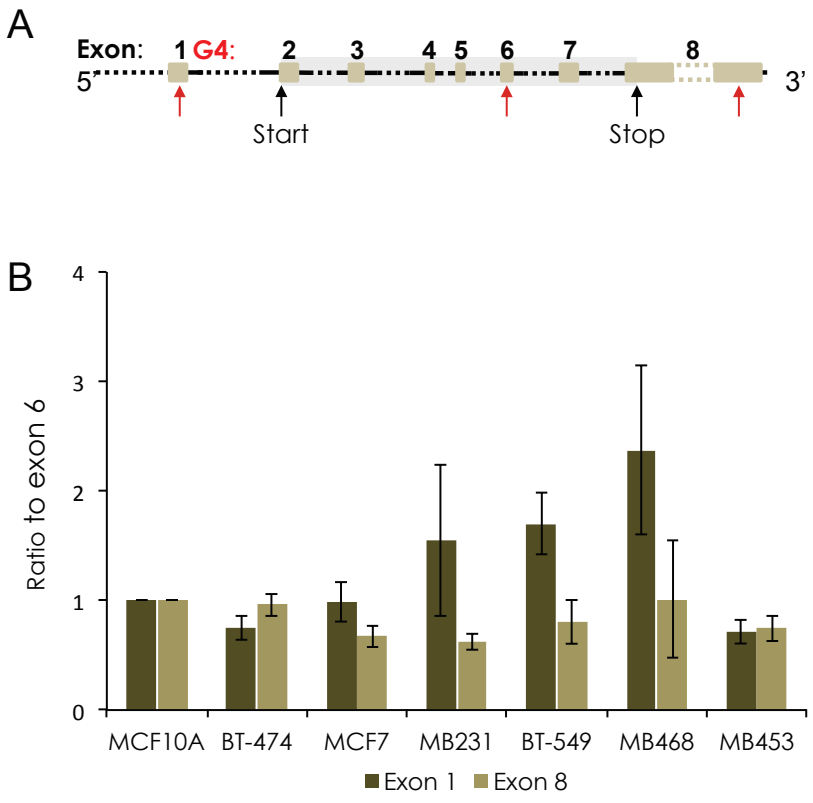

Figure 48: Expression analysis of the first exon of POLR3G. (A) Schematic view of POLR3G, the red arrows indicate the primer positions. (B) Using RT-qPCR analysis the expression levels of exons 1, 6 and 8 were measured. The expression of exon 1 and exon 8 were divided by those of exon 6, in order to normalize all expression levels. In the three triple-negative breast cancer cell lines MDA-MB231, BT-549 and MDA-MB468 the normalized expression of exon 1 is higher than that of exon 8. In all other cell lines the expression levels of exon 1 and exon 8 are about the same.

levels of POLR3G. This could indicate that the cell can detect the overexpression of POLR3G and tries to downregulate it by stabilizing the G4 in the first intron. The G4 is a steric hinderance to transcription by Pol II and this leads to higher levels of exon 1. If this is the case, than the regulation *via* G-quadruplex functions only partly as the POLR3G levels remain still high in the three triple-negative breast cancer cell lines.

Another hypothesis would be that the G4 serves as anchor point for transcription enhancers. While some transcription is aborted due to the sterical hinderance, overall transcription is enhanced. Further research is necessary in order to elucidate the possible function of the G-quadruplex in the first intron of POLR3G.

Chapter IV

Discussion

In 2010 Haurie *et al.* first described the existence of two different forms of RNA polymerase III subunit RPC32. Since then different studies have tried to answer the question of their respective function (Haurie *et al.*, 2010; Wong *et al.*, 2011; Renaud *et al.*, 2014; Lund *et al.*, 2017). The present study is the first to analyze the role of RPC32α in a tumor model.

Using bioinformatics, biochemistry and molecular biology, it was established that RPC32α is overexpressed in triple-negative breast cancer. This overexpression is not accompanied by an equivalent overexpression of other Pol III subunits, indicating that it was not a general upregulation of Pol III that lead to the high levels of RPC32α. Using CRISPR-Cas9, RPC32α knock-out cell lines were created from the MDA-MB231 mother cell line. This was the first time true RPC32α knock-out cells had been generated, all previous studies had worked with knock-down cell lines.

It was shown that no feed-back loop exists to regulate the expression of RPC32α. Furthermore, it could be demonstrated that RPC32α and RPC32β are regulated independently and the loss of RPC32α does not induce a rise in the expression of RPC32β. Similarly, other subunits of Pol III were unaffected by the loss of RPC32α, indicating that global Pol III levels remained unchanged. Thus the only possibly Pol III form in the RPC32α knock-out cells was Pol IIIβ.

While the knock-out cells presented no phenotypical difference, they formed significantly less colonies in soft-agar assays, indicating that RPC32α is important for tumorigenic growth. This result was confirmed by an *in vivo* experiment in mice. Cells lacking RPC32α showed a delayed onset of tumor growth and most importantly a 100 fold reduction of metastases. While these results give rise to the hope that RPC32α could become a new drug target in the fight against triple-negative breast cancer, a number of open questions remain.

1. What role for RPC32α and RPC32β?

The two forms of RPC32 show very distinct expression patterns. While RPC32α is highly expressed in tumor and stem cells, RPC32β is present in all somatic cells (Haurie *et al.*, 2010; Wong *et al.*, 2011). Early on these different expression profiles posed the question of cause and consequence. Wong *et al.* could show that a loss of RPC32α in stem cells lead to the differentiation of cells. Overexpression of RPC32α on the other hand protected stem cells from induced differentiation. Therefore RPC32α was important for the maintenance of stem-cellness in embryonic stem cells.

The present study analyzes the role of RPC32α in a tumor model. It could be shown that RPC32α is important for tumorigenic growth, which means the overexpression of RPC32α is one of the causes of tumorigenesis, not a mere consequence. The fact that both in stem and tumor cells the loss of RPC32α leads to such dramatic changes, can be explained in two different ways. Either RPC32α has a unique function and consequently RPC32β, which is still present in RPC32α knock-out cells, cannot

substitute it. Or RPC32β has the same function as RPC32α, but its expression levels are so low that it cannot rescue the RPC32α knock-out.

To test which hypothesis is correct, one would need to ectopically overexpress RPC32β in RPC32α knock-out cells and see if these new cells would be able to form colonies in soft agar. Haurie *et al.* (2010) showed that ectopic expression of RPC32β in transformed fibroblasts slightly inhibited colony formation in soft agar assays. Furthermore, ectopic expression of RPC32β leads to a moderate increase in the levels of tumor suppressor p53, while ectopic expression of RPC32α lead to a decrease of p53. Therefore it seems unlikely that the loss of tumorigenicity observed in RPC32α knock-out cells could be saved by an overexpression of RPC32β.

If an overexpression of RPC32β cannot save an RPC32α knock-out, the two forms would indeed have different functions. One of the first possibilities that comes to mind is the different selectivity for Pol III transcripts. This option is all the more likely, as RPC32α is part of a heterotrimer subcomplex which is involved in transcription initiation (Wang and Roeder, 1997). A previous study tried to identify possible difference in the transcription of Pol IIIα and Pol IIIβ, by chromatin-immunoprecipitation-sequencing (ChIP-seq) on fibroblasts (Renaud *et al.*, 2014). Four antibodies were used to identify potential Pol III transcription sites. The antibodies were directed against RPC32α, RPC32β, RPC53 and the TFIIIB subunit BDP1.

The authors identified a total of 293 loci occupied by either RPC32α and/or RPC32β. Both forms occupied genes of the three different promoter types (figure 49). Furthermore, both paralogs were

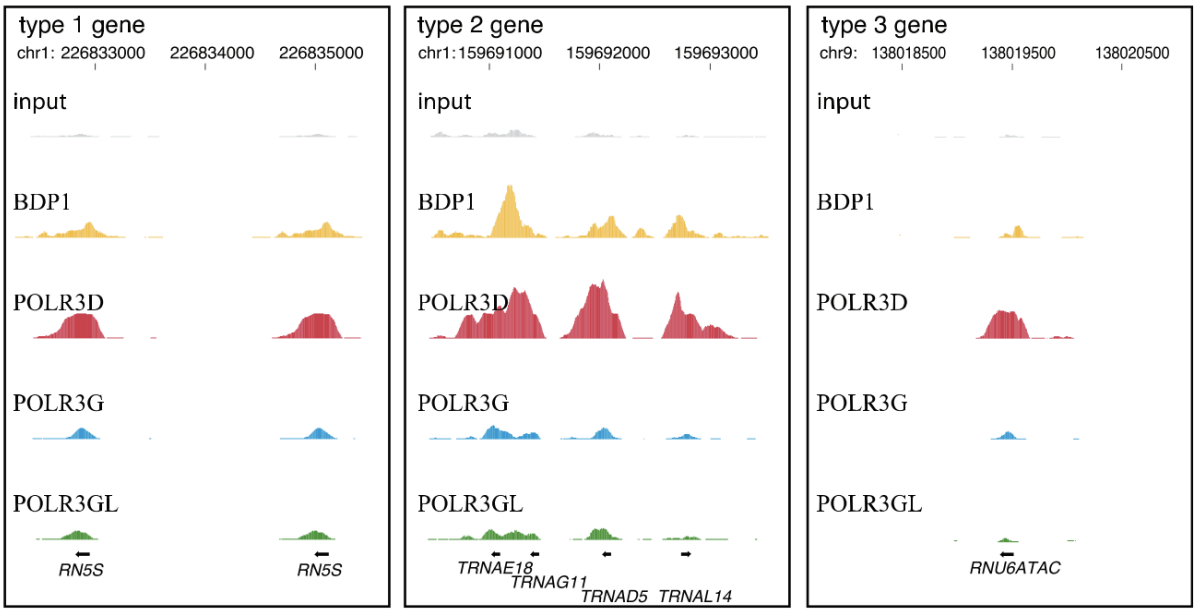

Figure 49: RPC32α and RPC32β occupy largely identical sites in IMR90 fibroblasts. The UCSC browser view shows the occupancy of RPC32α (POLR3G), RPC32β (POLR3GL), RPC53 (POLR3D) and TFIIIB subunit BDP1 on genes from the three different promoter types. The two RPC32 paralogs are found on all three promoter types and show very similar occupancy patterns.

Image from Renaud *et al.*, 2014

found on the large majority of Pol III transcribed genes, with the proportion for each form being almost constant. Among the few genes that were predominantly occupied by RPC32α were the U6 snRNA, a vault RNA and 17 tRNAs. The authors concluded that their data did not offer support for the hypothesis that the two forms have alternative sets of transcripts.

Two main arguments can be brought forth against this conclusion. First any ChIP-seq analysis depends largely on the sensitivity of the antibody. It is possible that a weak antibody might not detect rare sites of transcription. Second mere occupation of Pol III on a gene does not equal active transcription. If both RPC32α and RPC32β have been identified on Pol III transcribed genes, this does not indicate to what amount the two actively participate in their transcription.

Two other studies have previously analyzed the effect of a down-regulation of RPC32α on Pol III transcripts. Haurie *et al.* (2010) first identified the two RPC32 paralogs and Wong *et al.* (2011) characterized RPC32α in stem cells. The first study used siRNA to reduce levels of RPC32α, but makes no mention of any effect observed on Pol III transcripts. However, the authors note that upon ectopic expression of RPC32α in transformed fibroblasts they observed a strong augmentation of 5S rRNA, U6 snRNA, 7SK RNA and moderate increases in $tRNA^{Met\ i}$ as well as in RNA BC200. No such change was described for ectopic expression of RPC32β. Wong *et al.* (2011) use shRNA to induce an RPC32α knock-down in stem cells, which results in a strong overexpression of $tRNA^{Leu}$ and 5S rRNA and a down regulation of 7SL RNA. No effect was observed following the ectopic expression of RPC32α.

The observations made by the two studies are not easily reconciled. One describes a rise of 5S rRNA levels upon knock-down of RPC32α, the other upon overexpression of RPC32α. One study sees no alteration in Pol III transcript levels upon ectopic expression, the other relates strong increases for several transcript levels. An explanation might be the different nature of the cell lines used. A knock-down of RPC32α in transformed fibroblasts might not have the same effects as in stem cells. This underlines the importance of identifying a valid model to study the effect of RPC32α in tumors.

In the present study RPC32α was characterized using different breast cancer cell lines. The relevance of this model was shown by bioinformatic analysis of transcriptomic studies on clinical breast cancer samples. Using CRISPR-Cas9, a RPC32α knock-out cell line was created, which could be compared to its mother cell line. Comparison of the two cell lines revealed that the loss of RPC32α leads to a downregulation of some but not all Pol III transcripts. This is a strong indicator that Pol IIIα and Pol IIIβ transcribe each a different set of transcripts. However the number of transcripts analyzed *via* RT-qPCR was limited. To get a broader insight into the expression levels of different Pol III transcripts an RNA-seq analysis would have to be performed using RNA from the RPC32α knock-out cell line and the MDA-MB231 mother cell line.

2. How could Pol III transcripts promote tumorigenic growth?

2.1. tRNAs

If the hypothesis of differing sets of transcripts for Pol IIIα and Pol IIIβ is correct, than a certain set of Pol III transcripts has to promote tumorigenic growth. In this study the transcripts that were most affected by a loss of RPC32α were the $tRNAs^{Met\ i}$ and $tRNA^{Met\ e}$. Recently a study showed that the telomerase reverse transcriptase (TERT) can associate with RPC32α and that this association promotes tRNA transcription (Khattar *et al.*, 2016).

TERT is often found to be reactivated in cancers. For a long time it was believed that TERT was necessary to maintain telomere length in the rapidly proliferating cancer cells. However, several studies report non-canonical roles for TERT that are independent of telomere length (reviewed in Low and Tergaonkar, 2013 and Li and Tergaonkar, 2014). Using ChIP-seq analysis, Khattar *et al.* (2016) found that TERT localizes often in intragenic regions, 60% of which were genes transcribed by RNA polymerase III. The authors showed that TERT was significantly enriched at tRNA sites, but the enrichment differed between tRNA species. Interestingly the TERT enrichment depended on Pol III occupancy.

Via co-immunoprecipitation it was established that TERT interacted directly with RPC32α. TERT seems to direct RPC32α to specific chromatin sites, as cells that had high levels of chromatin-bound TERT also showed high levels of chromatin-bound RPC32α (figure 50 A). Intriguingly the overall level of RPC32α, as shown by western blot, did not change in the different cell lines (figure 50 B). This indicates that its recruitment to a genomic location is not dependent on the amount of protein, but rather on the presence of TERT.

The authors also showed that cells with low TERT levels could be rescued by overexpression of ectopic RPC32α. Furthermore, it was demonstrated in an *in vivo* experiment that loss of TERT lead to reduced mammary tumorigenesis in PyMT mice. These mice also showed reduced tRNA levels. These results led the authors to conclude that TERT associates with RPC32α, which promotes tRNA transcription. High levels of tRNAs would in turn favor tumorigenesis (figure 50 C).

Indeed the association of TERT with RPC32α is of interest, as the two proteins show similar expression patterns. Both are present at high levels in stem cells and tumor cells, but they have low expression levels in somatic cells. It would be interesting to analyze TERT chromatin binding patterns in the RPC32α knock-out cell lines. Does it still bind to the same sites as in the mother cell line or is the presence of RPC32α important to target TERT to specific regions? The authors did not test the capacity of TERT to bind RPC32β, but given that RPC32β is mainly expressed in somatic cells, the association seems less pertinent. The authors of the study also analyzed the correlation between

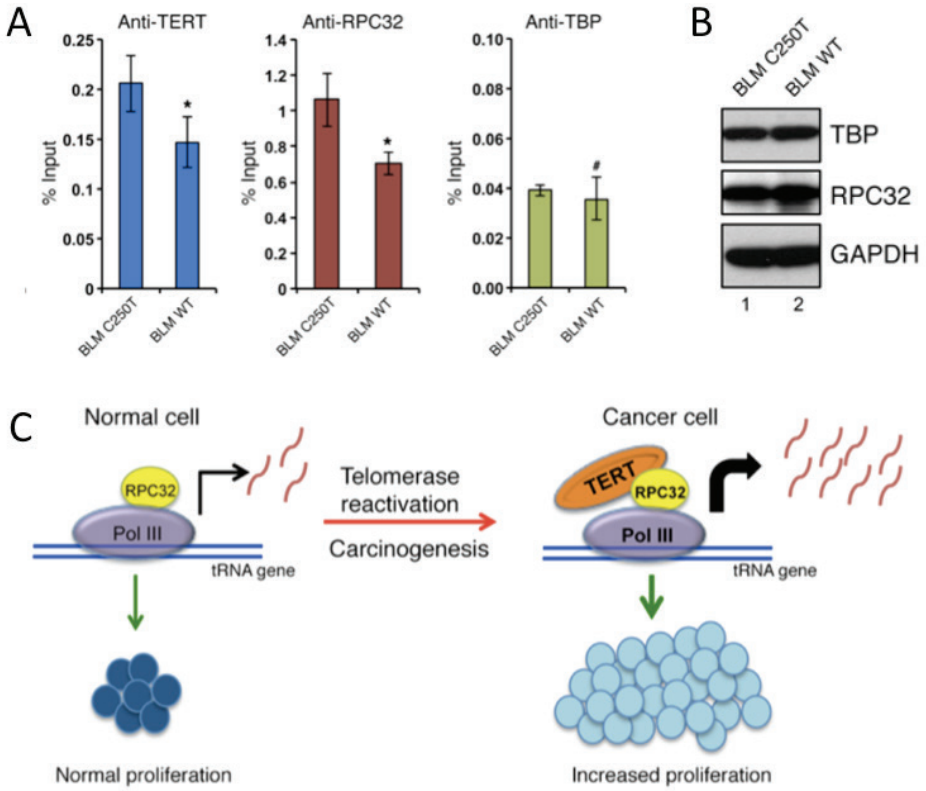

Figure 50: TERT increases RPC32α binding to the target genomic regions. (A) Occupancy levels of TERT (blue), RPC32α (red) and TBP (green) on the $tRNA^{Arg}$ genomic region. The BLM C250T cell line is a TERT mutant and shows higher TERT occupancy than the BLM WT line. This loss of occupancy is mirrored by RPC32α, but not by the TFIIIB subunit TBP. (B) In both BLM cell lines the levels of RPC32α protein are identical. (C) Proposed model in which TERT will bind to RPC32α. This will enhance tRNA transcription and lead to increased proliferation.

Image adapted from Khattar *et al.*, 2016

expression patterns of TERT and tRNAs in breast and liver cancers and found a positive and significant correlation in triple-negative breast cancers. A cancer type in which RPC32α is highly overexpressed, as demonstrated by the present study.

Khattar *et al.* (2016) could show an interaction of TERT and RPC32α, which resulted in higher tRNA levels. However, their study leaves several questions unanswered. Is TERT overexpression a cause or a consequence of carcinogenesis? How does the interaction of TERT and RPC32α promote tRNA transcription? Is RPC32α recruited as part of the Pol III or does it bind to TERT alone? Does TERT act as a transcription factor for Pol IIIα? They note that not all tRNAs are affected equally by a loss of TERT, but what tRNAs are needed for tumorigenic growth?

It would be interesting to analyze the TERT levels in the RPC32α KO cells. If the TERT levels are about the same as in the MDA-MB231 mother cell line, then high TERT would be a consequence not a cause of tumorigenicity. If the TERT levels have gone down in the KO cell lines, it should be tested if an overexpression of TERT in RPC32α knock-out cells can rescue the tumorigenic phenotype. This would indicate whether TERT can activate cancer growth independent of RPC32α or not. Furthermore, if an RNA-seq would be performed in the mother cell line and the RPC32α KO cell line, differentially

expressed tRNAs could be compared to tRNAs found highly expressed in TERT overexpressing cells. This could bring insights into which tRNAs are needed for tumorigenic growth.

Previous studies have shown that cells possess two different pools of tRNAs to either promote differentiation or proliferation (Gingold *et al.*, 2014). Furthermore, it was shown that the expression patterns of tRNAs vary in breast cancer cell lines compared to normal cell lines. In breast cancer cell lines tRNAs are found to be elevated 2 – 3 fold compared to the non-tumorigenic cell line MCF-10A (Pavon-Eternod *et al.*, 2009). Given that tRNAs make up for around 30% of total RNA, a two-fold increase is a substantial increase in transcription. The authors report that the increase in tRNA expression is not merely due to an increase in proliferation. Moreover the tRNA levels have not only increased, but the relative expression levels of different tRNA species have changed.

Moreover the authors noted changes in the use of tRNA isoacceptors. For example in MDA-MB231 cells the $tRNA^{Ser\ CGA}$ was found to be overexpressed, while the $tRNAs^{Ser\ TCA}$ shows levels similar to the MCF-10A control cell line. The authors concluded that tRNA isoacceptor expression was optimized for the expression of cancer related proteins, which favor tumorigenesis. In a follow up study, it was shown that the overexpression of $tRNA^{Met\ i}$ was sufficient to reprogram tRNA expression levels and to increase proliferation in human epithelial cells (Pavon-Eternod *et al.*, 2013). In a more recent study two tRNAs ($tRNA^{Glu\ UUC}$ and $tRNA^{Arg\ CCG}$) were identified that actively promote breast cancer metastases (Goodarzi *et al.*, 2016).

While this is evidence that tRNA isoacceptors play a role in modifying cellular functions, there might be another level of regulation. Different genes coding for the same tRNA, but with varying sequences are known as isodecoders. More than 50% of human tRNA genes are isodecoders (Goodenbour and Pan, 2006). While the sequence variations do not have an effect on the secondary or tertiary structure, it was shown that the isodecoders can have functional variations (Geslain and Pan, 2010). Some isodecoders may even perform functions distinct from translation (Geslain and Pan, 2010).

TRNAs are just one group of transcripts synthesized by Pol III, but they alone are already potent enough to deregulate cellular functions and promote tumor growth. It is therefore possible that RPC32α favors tumorigenesis by initiating transcription of a certain subset of tRNA isoacceptors or isodecoders, which will push the cell into tumorigenesis. It would be of great interest to analyze the tRNA isoacceptor and isodecoder levels in the RPC32α knock-out cell line created in the present study. Comparing the levels in the knock-out cell line with the mother cell line, would help to identify distinct tRNAs that favor tumorigenesis.

2.2. tRNA derived fragments (tRF)

Another level of regulation comes from tRNA derived fragments (tRF) (figure 51 A and B). These small RNAs may serve as interfering RNA, translational regulators or epigenetic factors. Initially they were discovered through bioinformatic filtering of deep sequencing data (Cole *et al.*, 2009). Since their identification, it was revealed that they are implicated in a variety of diseases, notably cancer (reviewed in Soares and Santos, 2017) (figure 51 B). For example it was found that breast cancer patients had in their blood specific sets of 5'tRNA-halves (Dhahbi *et al.*, 2014).

Another study found a functional link between tRFs and metastatic breast cancer (Goodarzi *et al.*, 2015). The authors showed that a specific set of tRFs will bind to the RNA binding Y-box protein 1 (YBX1). This protein has been shown to be implicated in a variety of cancers (reviewed in Matsumoto and Bay, 2005). The Y-box protein family is involved in different steps of the mRNA metabolism by regulating transcription, splicing, mRNA stability and translation. Goodarzi *et al*. (2015) were able to

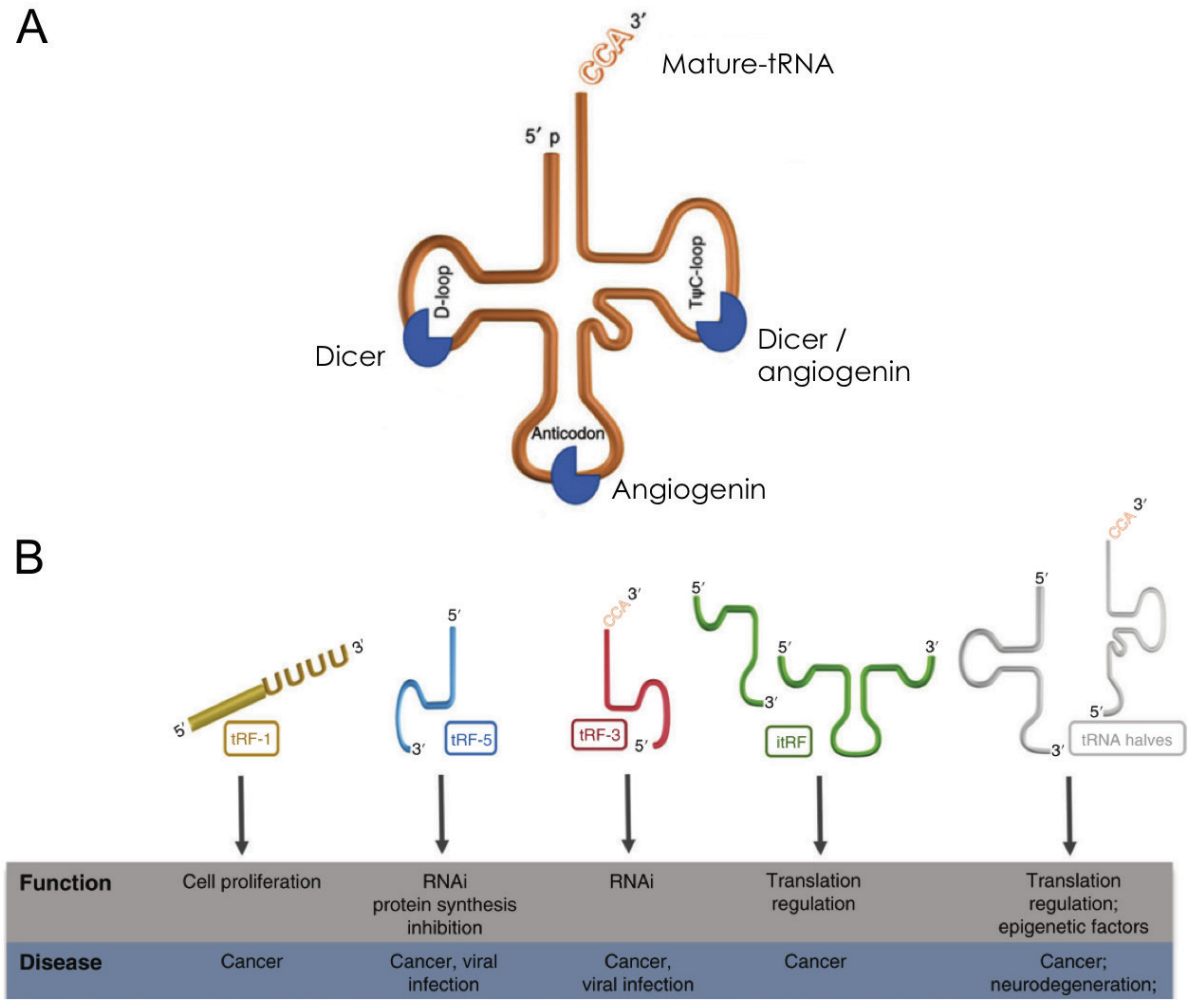

Figure 51: Different types of tRNA derived fragments (tRF) and their implications in cancer. (A) Structure of a mature tRNA. The blue pacman symbols indicate potential cutting sites and the enzymes responsible. (B) The resulting tRFs and their identified implications. The yellow tRF stems from the 3' end of a pre-tRNA.

Images adapted from Soares and Santos, 2017

demonstrate that YBX1 stabilizes oncogenic transcripts and thereby promotes tumorigenic growth. However, tRFs can counteract this mechanism by binding to YBX1 and competitively displacing other YBX1 bound transcripts. Thereby oncogenic transcripts are no longer stabilized by YBX1 and will be degraded. This in turn suppresses metastatic progression. Interestingly these protective tRFs were downregulated in highly metastatic breast cancer cell lines upon hypoxia.

These studies show that tRFs are active regulators of cell function. It is possible that Pol IIIα transcribes a different set of tRNAs, which will then be cleaved into tRFs. The presence or absence of distinct tRFs might lead to a deregulation of tumor suppressors and or oncogenes, which will favor the onset of tumorigenic growth. Therefore it would be interesting to analyze the levels of tRFs in the RPC32α KO cell lines and compare them to those of the mother cell line. This might help identify tRFs that are transcribed by Pol IIIα, but not Pol IIIβ and that are implicated in tumorigenesis.

2.3. Other Pol III transcripts

In the present study $tRNA^{Met\ i}$ and $tRNA^{Met\ e}$ were the Pol III transcript with the most significant reduction upon RPC32α knock-out. Another transcript that showed a significant downregulation was 7SL RNA. A previous study had shown that high levels of 7SL RNA repressed activity of p53 (Abdelmohsen *et al.*, 2014). The tumor suppressor p53 negatively regulates transcription by Pol III (Zambetti *et al.*, 1992). This leads to the hypothesis that high levels of RPC32α generate high amounts of 7SL, which in turn will suppress p53. Low levels of p53 would lead to augmented Pol III transcription, which include higher levels of 7SL *etc.*

The Pol III transcriptome is diverse and probably not yet identified completely. In this study only a few transcripts have been analyzed *via* RT-qPCR. To get a broader understanding of what transcripts are deregulated, an RNA-seq analysis would be needed. Recently a study examined RPC32α depended transcription in human embryonic stem cells (hESCs) (Lund *et al.*, 2017). The authors of the study compared normal hESCs and hESCs in which RPC32α had been knock-down using siRNA. Deep-sequencing analysis was performed to analyze both small RNAs and mRNAs.

The class of small RNAs that showed the most changes was mature miRNA (n=59), representing 61% of all small RNAs affected by the RPC32α knock-down. In second and third place came pre-miRNAs (n=20) and snoRNAs (n=10), corresponding to 21% and 10% of all RPC32α dependent small RNAs respectively. These results are surprising, as none of the most affected small RNAs are transcribed by Pol III. The only Pol III transcripts that were significantly deregulated were three 5S rRNA pseudogenes and two tRNAs ($tRNA^{Thr\ ACY}$ and $tRNA^{Leu\ TTA}$). The absence of Pol III transcripts is partly due to a methodological bias. The authors used very strict statistical filters, which is necessary in case

of multiple testing. It is possible that some deregulated Pol III transcripts were filtered out, even though their transcription levels had changed.

Another bias is that the analysis looks at the fold change of each RNA. However, a twofold change in an abundant type of RNA, such as tRNAs, corresponds to vast quantities of additional RNA. To achieve a twofold change in a less abundant RNA only small amounts of extra RNA are need. It is possible that some of the Pol III transcripts were deregulated, but their fold change did not make the cut in the significance analysis. Lastly, it is not because a fold change is not statistically significant, that it cannot induce change in cellular processes. Especially regulatory RNAs, like many of the Pol III transcripts, might produce a big effect, with only a small change.

While the absence of Pol III transcripts might in part be explained by these biases, it is still worth noting that it is small RNAs transcribed by Pol II and not Pol III that are most affected by a knock-down of RPC32α. The majority (n=84, 87%) of the small RNAs that showed altered expression levels were downregulated. Only 13% (n=13) of all small RNAs were upregulated (figure 52). This shows that in stem cells, RPC32α is necessary for the maintenance of transcription of small RNAs and it only has a limited suppressive function.

The same study also analyzed the changes in mRNA transcription upon RPC32α knock-down (Lund *et al.*, 2017). The most abundant RNA class to be altered upon RPC32α knock-down were protein coding mRNAs (n=593), which represents 83% of all deregulated transcripts and 5% of all protein coding genes in the hESCs transcriptome. Like the small RNAs the majority of transcripts (n=681; 94,8%) was downregulated rather than upregulated (n=37; 5,2%) (figure 52). Again RPC32α seems to have an activating rather than a suppressing function.

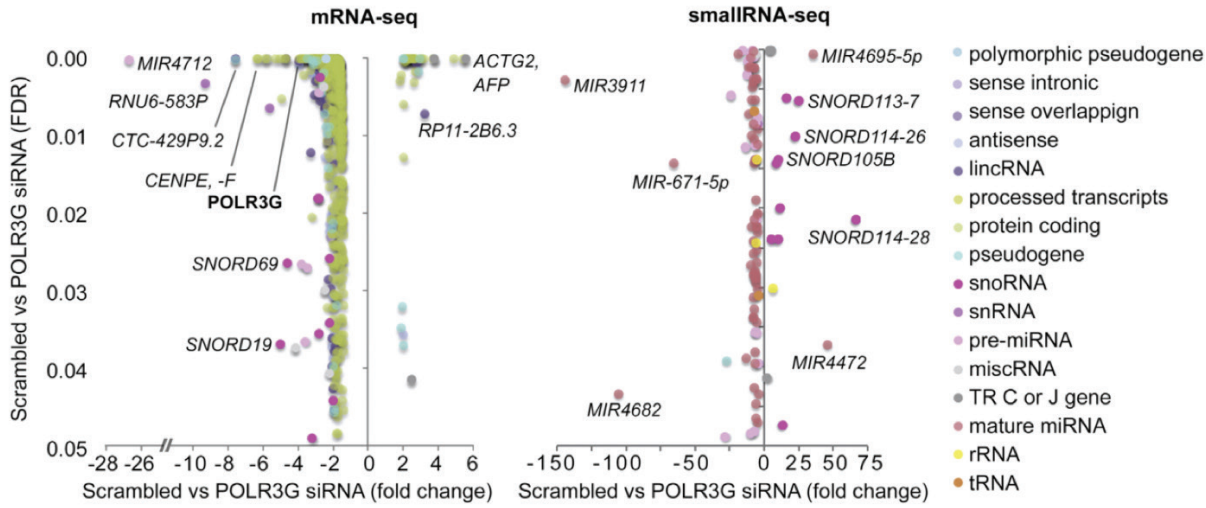

Figure 52: RNAs altered after the knock-down of RPC32α in human embryonic stem cells. On the left hand side are all the PolyA$^+$ RNA transcripts altered after RPC32α knock-down. The majority of them being protein coding mRNAs (green). On the right hand side the small RNAs that showed changed transcription levels upon RPC32α knock-down. In both RNA-seq analyses the majority of transcripts is downregulated.

Image from Lund *et al.*, 2017

Interestingly the mRNA of RPC32β was not upregulated upon RPC32α knock-down. This is in line with the results presented in the present study. The breast cancer RPC32α knock-out cells also did not show an upregulation of RPC32β, neither on an RNA nor protein level. This indicates that while the two paralogs have reverse expression patterns, they are not co-regulated.

The results of this study show that a loss of RPC32α has a profound impact on the transcriptome of a cell and the changes in transcription reach far beyond Pol III transcripts. One possibility of how Pol IIIα can influence even mRNA levels, is by favoring transcription of distinct regulatory RNAs. Another possibility would be that RPC32α and/or Pol IIIα have functions other than transcription.

It would be of great interest to perform a similar transcriptomic analysis with the RPC32α cell line and the mother cell line. The direct comparison of two cell lines, which have the same genetic background and differ only in the expression of RPC32α, would reveal what transcripts are directly or indirectly regulated by RPC32α. Those results would help further the understanding of tumor onset. Furthermore, a comparison of such a transcriptomic analysis to the results of the study done in stem cells (Lund *et al.*, 2017), might reveal new mechanisms involved in the creation and maintenance of tumor stem cells.

2.4. Roles for RPC32α beyond transcription

The results obtained by the present study, show that RPC32α plays an important role in tumor initiation. But the exact mechanisms at work remain to be elucidated. As RPC32α is part of an RNA polymerase, a role in transcription seems plausible. However, it is also possible that RPC32α has taken up other functions in the cell, either as part of the polymerase or independently. Indeed signs exist that point to a larger role for Pol III than mere transcription.

Several ChIP-seq studies have tried to map Pol III to the genome, in order to identify new Pol III transcripts. One surprising finding was that in HeLa cells only 52% of the *in silico* predicted tRNAs were occupied by Pol III (Oler *et al.*, 2010). This was in contrast to yeast cells where almost all tRNAs had been found occupied and transcribed by Pol III (Harismendy *et al.*, 2003). Often the occupied tRNAs lay just upstream of Pol II genes, whereas unoccupied tRNAs did not cluster near Pol II genes (Oler *et al.*, 2010). Most of the occupied tRNA genes gave two peaks at -300 and -900 nucleotides of a Pol II transcription start site (TSS), revealing a common tandem gene structure. Typically the tRNAs, which are located to Pol II TSS are transcribed away from the Pol II gene, so that the two polymerases will not interfere with each other (Oler *et al.*, 2010).

Peaks of Pol III occupancy often were in close proximity to Pol II occupancy (Raha *et al.*, 2010). Interestingly the occupancy of Pol II and Pol III coincided not only at annotated Pol II promoters, but also outside of them (figure 53). In HeLa cells as much as 82% of occupied tRNAs were found outside

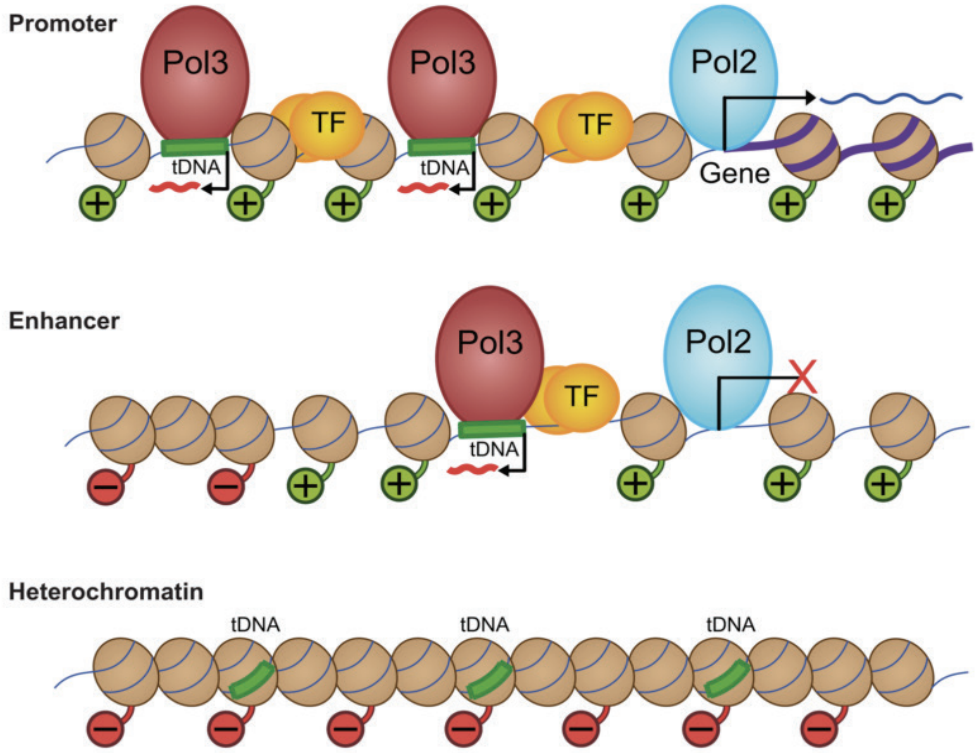

Figure 53: Chromatin opening and the relationship between Pol II and Pol III. In annotated Pol II promoters tandem tRNA genes were found to be occupied by Pol III, with their transcription being directed away from the Pol II gene. Occupied tDNA has also been found in non-annotated Pol II genes, but there as well Pol III presence coincided with occupancy of Pol III, Pol II transcription factors (yellow, TF) and open chromatin (green + signs). TDNA that was not occupied by Pol III was mostly associated with closed chromatin (red – sign).

Image from Oler *et al.*, 2010

of annotated Pol II promoters. Also Pol III occupancy correlated with histone marks associated with Pol II transcription start sites and occupied tRNA coincided with open chromatin (Moqtaderi *et al.*, 2010; Oler *et al.*, 2010).

These results reveal a close interaction between Pol II and Pol III, but the functional relationship between the two is still unclear. It has been suggested that the chromatin is made accessible by factors related to Pol II and that Pol III simply benefits from this opening. Indeed the inhibition of Pol II with α-Amanitin lead to a downregulation of a subset of Pol III transcripts (Raha *et al.*, 2010). However, there are some signs that Pol III can actively contribute to chromatin opening, in which case it would be Pol II that benefited from the opened chromatin.

Almost two decades ago, it was shown that the Pol III transcription factor TFIIIC has an intrinsic histone acetyltransferase activity that can relieve chromatin mediated repression of Pol III transcription (Hsieh *et al.*, 1999; Kundu *et al.*, 1999). Furthermore, it was shown that TFIIIC can bind the histone acetyltransferase p300 (Mertens and Roeder, 2008). While these findings show that Pol III can actively influence chromatin opening, how does this affect Pol II transcription? And does RPC32α play a role in chromatin opening?

In a combined ChIP-seq transcriptome experiment in hESCs, 225 sites were identified that were occupied by two subunits of Pol III (RPC160 and RPC32α). These sites were further analyzed for transcriptomic changes in RPC32α knock-down cells. One mRNA was identified that showed a promoter proximal binding of Pol III and that was downregulated upon RPC32α knock-down (Lund *et al.*, 2017). The mRNA in question codes for POLG a subunit of the mitochondrial DNA polymerase. With only one mRNA that seems to be directly regulated by Pol III, the effect of Pol III occupancy at Pol II promoters seems to be limited. However, 225 sites analyzed in the whole genome may be not fully representative of all the Pol III and Pol II interactions that might occur. The fact that at least one mRNA has been identified, which seems to be regulated by Pol III, is prove that the interactions between Pol II and Pol III can go both ways.

A more striking finding of the ChIP-seq transcriptome study was that numerous tRNA genes were bound by RPC32α, but upon RPC32α knock-down these tRNAs did not show transcriptomic alterations. This indicates that the RPC32α bound to the tRNAs was not indicative of their transcription. This poses the question of the role of RPC32α at those tDNA sites. These findings point towards a role for RPC32α and/or Pol IIIα outside of transcription and potentially in chromatin regulation.

To further elucidate the question of a potential role of RPC32α outside transcription a ChIP-seq analysis could be performed. Using antibodies specific to certain histone modifications, such as acetylation or methylation, the RPC32α knock-out cell line and the mother cell line could be analyzed. Comparing the results of the two cell lines would show if the presence of RPC32α induces Pol III to bind at sites with a distinct histone profile. This would indicate that RPC32α either recognizes a distinct epigenetic profile or that it can create a certain epigenetic environment that favors binding of Pol III. The question of a potential role for RPC32α independent of Pol III could be analyzed *via* immunoprecipitation (IP). If RPC32α is used as bait, proteins binding to RPC32α could be purified and analyzed by western blot of mass spectrometry. If proteins other than the Pol III transcription apparatus are purified with RPC32α, this could indicate a possible role for RPC32α independent of Pol III. An interesting negative control would be to use the newly identified protein as bait in a cell extract of the RPC32α knock-out cell line. If the interaction with the protein is RPC32α specific, than the protein should not co-precipitate with proteins from the Pol III transcription apparatus. However, if the protein does attract other Pol III subunits or transcription factors even in the absence of RPC32α, than the interaction is not dependent on RPC32α. While these experiments would not prove that RPC32α acts independently of Pol III, it could reveal new specific interaction partners that would give insight into what role RPC32α might play outside of transcription.

3. How is RPC32α regulated?

3.1. G-quadruplexes

RPC32α is found only in distinct types of cells, while it is suppressed in others. Therefore the expression of POLR3G needs to be closely regulated. Previously it had been discovered that the promoter region of POLR3G contains a MYC binding site and upon the induction of ectopic MYC there is accumulation of MYC on the POLR3G TSS (Renaud *et al.*, 2014). No MYC binding site was found in the promoter region of PORL3GL, indicating a differential regulation of the two subunits. Given the oncogenic nature of POLR3G it seemed likely that other regulatory mechanisms exist, to control the expression of POLR3G.

In the present work a consensus G-quarduplex (G4) sequence was identified in the first intron of POLR3G, but not in POLR3GL. The sequence was shown to form a G-quadruplex *in vitro* and the structure proved to be stable at temperatures beyond the average body temperature. This led to the hypothesis that the G4 might also form *in cellulo*, where it possibly serves a regulatory function.

G-quadruplexes have a consensus sequence of $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$. The guanines are able to interact *via* Hoogsteen hydrogen bonding and form so called guanine tetrads. The tetrads are often stabilized by a cation, mostly potassium. Several tetrads can stack up to form a G-quadruplex (figure 44 A). Almost three decades ago a study showed that telomere sequences at the end of each chromosome where able to form G4 *in vitro* (Sundquist and Klug, 1989). For a long time it was thought that the role of G-quadruplexes was limited to protecting the telomeres from degradation.

Nonetheless early studies already reported on the presence of G4 sequences in other regions of the genome. One study showed that a G4-quadruplex sequence was present in the control region of c-Myc (Simonsson *et al.*, 1998). First evidence that this G4 sequence might play a functional role was provided by a study in 2002. Using luciferase reporter assays, the authors showed that the G-quadruplex suppressed transcription (Siddiqui-Jain *et al.*, 2002). Furthermore, the authors showed that the G4 stabilizing drug TMPyP4 could lower *in vitro* transcription of c-Myc in a G4 sequence specific manner. The isomer TMPyP2, which has a low affinity for G4, had no affect on c-MYC transcription. Another study showed that TMPyP4 was able to downregulate c-MYC *in cellulo,* while TPMyP2 had no effect (Grand *et al.*, 2002).

Since the discovery of the c-MYC G-quadruplex, other cancer related genes were shown to possess similar G4 sequences in their promoter. Using *in vitro* and *in vivo* methods it was shown that these sequences could form G-quadruplexes, which can impact transcription or DNA synthesis (figure 54 A). Some of the genes analyzed were the human vascular endothelial growth factor (VEGF) gene (Sun *et al.*, 2005); the retinoblastoma susceptibility gene (Rb) (Xu and Sugiyama, 2006); the Kirsten

ras oncogene homolog (KRAS) (Cogoi and Xodo, 2006; Cogoi *et al.*, 2008) and the VEGF receptor 2 (VEGFR-2) (Salvati *et al.*, 2014).

Given that G4s are implicated in the regulation of varying oncogenes, they have become interesting targets for anti-cancer drugs. The G4-interactive-compound quarfloxin, directed against c-Myc, had moved all the way to phase II clinical trials, before it had to be abandoned due to problems with bioavailability (reviewed in Balasubramanian *et al.*, 2011). More recently the G4 stabilizer CX-5461 was shown to selectively inhibit tumor growth in BRCA1/2 deficient tumors in mice and is now in phase I safety testing in humans (Xu *et al.*, 2017).

While G-quadruplexes in promoters have gotten much attention, they are not the only potential regulatory G4 structures. A bioinformatic analysis revealed that G4 motifs are also a common feature in the first intron of the non-template strand (Eddy and Maizels, 2008) They are located on average 200 nt downstream of the TSS and within 100 nt of the 5' end of the first intron. The G-quadruplex identified in the present study belongs to this same group of G4s. It was shown that the G-quadruplexes in the first intron correlate with promoter proximal pausing of Pol II (Eddy *et al.*, 2011). It is therefore plausible that the G4 structure is a regulatory element in transcription. The G4 could be stabilized by ligands to repress transcription or a G4 helicase could be needed to enable transcription.

If the G4 does indeed block transcription than aborted transcripts corresponding to exon 1 should be present in the cell. In this study it was shown that exon 1 of POL3RG is overexpressed compared to the rest of the gene in the triple-negative breast cancer cell lines. All other cell lines showed equal transcription of exon 1 and exon 8. As the triple-negative cell lines are also the one that overexpress the gene as a whole, it could be possible that the cell detects this overexpression and tries to downregulate it by stabilizing the G4 (figure 54 B). However, transcription is so high that the blockage by the G4 potentially reduces the overexpression, but it cannot prevent it. To test for this hypothesis RPC32α would need to be overexpressed in one of the other cell lines, to test if a similar overexpression of exon 1 would then be observed.

Another explanation would be that the G-quadruplex serves as an anchor point for transcription enhancers (figure 54 C). While the G4 might still block transcription, the overall positive effect of the enhancers is such that the expression levels rise. If the G4 structure is resolved or if the polymerase succeeds to transcribe the gene in the presence of the G4, than the G rich sequence would be found again in the RNA (figure 54 D). Therefore a G-quadruplex could possibly be formed in the RNA and thus inhibit translation. Given the results obtained from G4 structures in promoter regions, it is highly plausible that G4s in introns can fulfill similar regulatory roles.

To further investigate the function of the G4 identified *in vitro* by the present study, it would be necessary to confirm its formation *in cellulo*. For this the G4 sequence would needed to be cloned in

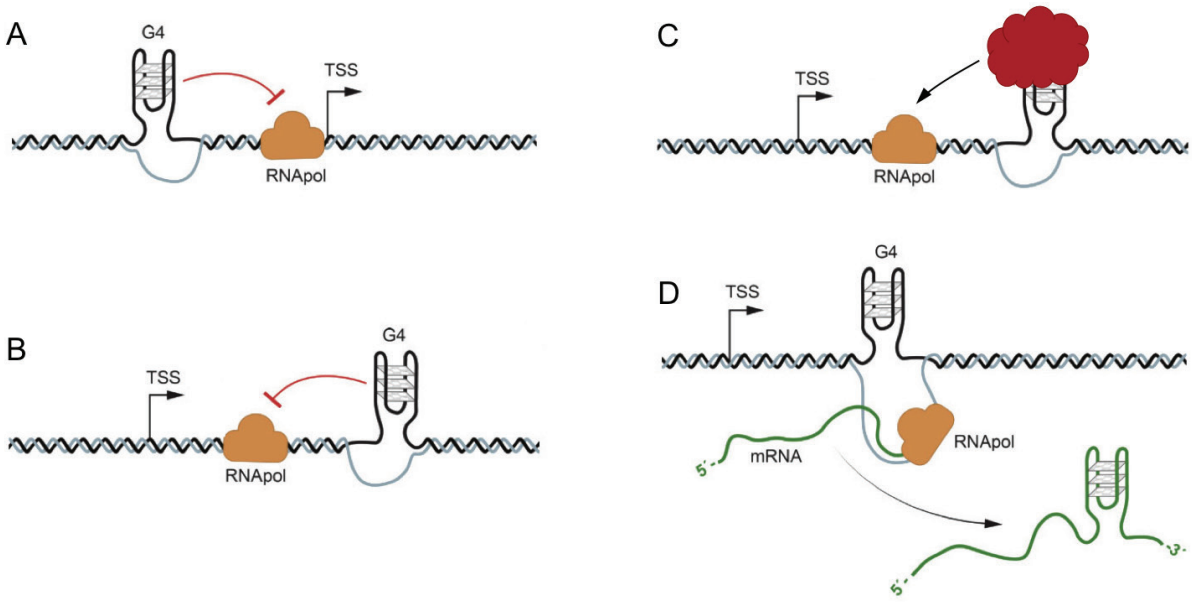

Figure 54: Potential effects of a G-quadruplex on transcription. (A) In promoters of several oncogenes, G-quadruplex have been shown to hinder transcription. (B) In this work a G-quadruplex was identified in the first intron of POLR3G, which might also affect transcription. Either the G4 could block transcription or (C) it could serve as anchor point for transcription enhancers. (D) If the RNA polymerase succeeds to transcribe the gene, a G-quadruplex might form in the RNA and hinder translation.

Images adapted from Bochman *et al.,*2012; Rhodes and Lipps, 2015

front of a luciferase reporter gene, which then is transfected into the cell. Luminescence could be measured before an after the addition of G4 stabilizers like TMPyP4. A change in luminescence would indicate that the G4 indeed affects transcription. A decrease would point to a negative regulatory role, whereas an increase would suggest that the G4 could attract transcriptional activators. The TMPyP4 isomer TMPyP2 could serve as a negative control, as it does not stabilize G4s, no effect should be visible upon addition of TMPyP2.

Furthermore, the overall effect of stabilizing the G4 in the first intron of POLR3G could be analyzed in the cells. Addition of TMPyP4 to the cells would stabilize the G4 and subsequently the cells be examined for potential changes in RNA or protein levels. To guard from potential off target effects of TMPyP4, the experiment could be repeated in the RPC32α knock-out cells. The G4 in question is still present in the cells, but theoretically stabilizing the G4 should not affect the cell as the protein is no longer translated. A change in transcription of POLR3G would go unnoticed. However potential off target effects would also be visible in the KO cell lines. Another interesting experiment would be to test if stabilizing the G4 leads to increased Pol II pausing at the beginning of POLR3G. This would point towards a negative regulatory role of the G4.

3.2. miRNAs

In the present study it has been suggested that the expression of POLR3G is regulated by a G-quadruplex in the first intron of the gene, but other regulatory mechanisms are possible. The expression of many mRNAs is regulated via miRNAs. Therefore it is possible that POLR3G too is regulated on an RNA level. Using bioinformatics POLR3G was identified as a potential target of miR-27 in hESC (Fuchs *et al.*, 2014) (figure 55). To verify the interaction between miR-27 and POLR3G, a plasmid containing the region with the predicted binding site of POLR3G was cloned into a GFP vector. Human embryonic kidney cells 293 (HEK293) were transfected with the plasmid and a miRNA mimic or negative control. Cells were analyzed via flowcytometry after 48 hours.

It was shown that the fluorescence of the plasmid containing the binding site from POLR3G was downregulated by 20% compared to the control. Besides POLR3G, miR-27 also downregulated the fluorescence of plasmids containing sequences from other pluripotency associated genes, such as NANOG and LIN28. The authors showed that miR-27 was upregulated when hESC underwent differentiation. Furthermore, knock-down of the pluripotency factor OCT4 lead also to an upregulation of miR-27. These observations are in line with POLR3G being highly expressed in stem cells, but downregulated upon differentiation.

Recently another miRNA was found to regulate POLR3G expression in human pluripotent cells. By comparing microarray data from hESC and differentiated cells, miR-1305 was identified as a potential regulator of pluripotency (Jin *et al.*, 2016). Using bioinformatic tools POLR3G was identified as a possible target of miR-1305. *In vitro* experiments showed that miR-1305 could induce a 45% downregulation of a POLR3G reporter gene, which was rescued when the miR-1305 binding site was mutated. The authors also showed that overexpression of miR-1305 induced differentiation in human induced pluripotent stem cells (hiPSC). But cells could be protected from differentiation by overexpression of POLR3G. Inversely knock-down of miR-1305 enhanced pluripotency, but the effect was lost when POLR3G was knocked-down as well.

In cancer cells the miRNA miR-223 was identified as a potential regulator of POLR3G expression (Wu *et al.*, 2013) (figure 55). Cells had been transfected with a luciferase reporter plasmid containing the miR-223 target region in POLR3G. Transfection with miR-223 lead to drastically reduced levels of luminescence. To verify that the reduction was depended on miR-223, the authors transfected the cells with an additional plasmid containing one of several oligo decoys. The decoys contained sequences of variable complementarity to miR-223. It could be shown that the more the decoy was complimentary to miR-223, the less the luminescence of the POLR3G containing plasmid was reduced. Proving that the effect on POLR3G was miR-223 depended.

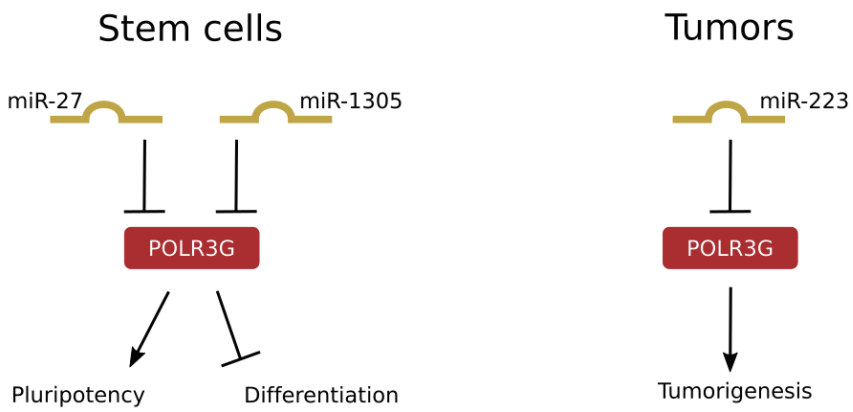

Figure 55: MiRNA regulating the expression of POLR3G. In stem cells the miRNAs miR-27 and miR-1305 were shown to negatively regulate POLR3G, which is a key factor for the maintenance of pluripotency. In tumor cells the miRNA miR-223 was identified as a regulator of PORL3G. The triple-negative cell line MDA-MB231 was found to have low levels of miR-223.

Another study showed that overexpression of miR-223 reduced cell migration and invasion in MDA-MB231 cells (Pinatel *et al.*, 2014). Furthermore, it was shown that non-tumorigenic MFC-10A cells have very high levels of miR-223, while luminal breast cancer line MCF7 has somewhat lower levels and triple-negative line MDA-MB231 has very low levels of miR-223 (Gong *et al.*, 2013). The miR-223 expression levels are therefore inversely proportional to the POLR3G observed by the present study. These results lead to the hypothesis that miR-223 negatively regulates POLR3G, the gene coding for RPC32α. As RPC32α is important for tumorigenic growth, high levels of miR-223, like in MCF-10A cells, lead to low levels of RPC32α and to protection from tumorigenic growth. On the other hand low levels of miR-223, as in MDA-MB231 cells, lead to high expression of RPC32α and aggressive tumorigenic growth.

During the course of the present study, the expression levels of miR-27 and miR-223 have been analyzed in the different breast cancer cell lines. However, no conclusive results could be obtained. Several factors could explain why the results obtained by previous studies could not be reproduced. One explanation could be that cell culture conditions were not the same as the ones described in the study. It is possible that miRNA levels were influenced by factors present in the serum used. However, between the different studies reporting low levels of miR-223 in MDA-MB231 cell lines, the cell cultures differed as well, yet they arrived at the same result (Gong *et al.*, 2013; Pinatel *et al.*, 2014). To test for the influence of culture conditions, the cells were kept in conditions as described in one study, but the cell line MCF-10A did not continue to grow under these conditions. Any results obtained under these conditions could therefore not be normalized to MCF-10A, which made inter-cell line comparisons difficult.

A bigger challenge was the identification of a suitable housekeeping gene for normalization. In the studies cited, the miRNA expression is normalized to the U6 snRNA expression. It is a standard used

in many miRNA studies. However, the U6 gene is transcribed by the RNA polymerase III. Given that RPC32α is part of the Pol III, it did not seem wise to analyze regulation of POLR3G by using a Pol III transcript for normalization. After all fluctuations on POLR3G might have repercussions on the transcription level of Pol III and on U6 snRNA levels. A lot of time was devoted to finding a potent replacement for U6 as normalizer and finally the small nucleolar RNA U48 was identified. It is possible that by normalizing to U48, the expression profiles were altered compared to normalization to U6. So while no conclusive results were obtained in the present study, it does not mean that POLR3G might not be regulated by the above mentioned miRNAs. Future studies might examine this possibility further.

4. Is RPC32α related to cancer stem cells?

In the present study it was shown that loss of RPC32α leads to small tumors and dramatically less metastases *in vivo*. Analysis of the primary tumor via hematoxylin and eosin (HE) staining showed that the tumor formed by wild type cells had invaded the surrounding tissue. Tumors formed by RPC32α knock-out cells also invaded the surrounding tissue, but to a lesser degree. The decreased invasiveness can most likely be explained by the smaller tumor size of the KO cell tumor.

Immunohistochemical staining did not reveal any alterations in the amount of estrogen receptors, confirming the triple-negative status of the KO tumor. Analysis of the antigen Ki67 was also found to be unaltered, indicating that proliferation levels were the same in wild type and KO tumors. One explication why no difference was observed is that immunohistochemical staining might not be sensitive enough. Other techniques such as western blotting or RT-qPCR might be more appropriate. As part of this study preliminary test using western blots have been performed for EMT markers, such as E-cadherin and vimentin. However, the quality of the western blots was not sufficient to draw any conclusions. Also the levels for some stem cell markers such as NANOG and OCT4 have been tested by RT-qPCR, but the expression levels in both mother cell line and KO were too weak, to obtain reliable results.

Another explication why immunohistochemical staining did not show differences between the two cell lines is that the bulk of the tumor cells were not affected by the loss of RPC32α and only the cancer stem cells had lost their tumorigenic potential. Given RPC32αs role in embryonic stem cells, namely to maintain proliferation, it is conceivable that RPC32α also is important for the maintenance of cancer stem cells. It would therefore be important to examine the mother cell line MDA-MB231 and the RPC32α knock-cell lines for the presence of known stem cell markers.

A well established stem cell marker is the overexpression of hyaluronic acid receptor CD44. In this study both wild type and KO tumors showed high levels of CD44, which seems to indicate that RPC32α is not implicated in the maintenance of stem-cellness in cancer, but again immunohistochemical staining might not be sensitive enough to detect differences. Also, while CSC are characterized by high levels of CD44, the presence of CD44 alone, does not identify a stem cell. In the original study that identified the $CD44^+/CD24^-$ signature for cancer stem cells in breast cancer, the authors showed that $CD44^+/CD24^+$ cells were unable to initiate tumor formation (Al-Hajj *et al.*, 2003). Therefore it would be necessary to test both the cell lines and tumors for the presence of CD24, before concluding on RPC32αs capacity to maintain stem-cellness.

In a more recent study it was shown that the level of $CD44^+/CD24^-$ cells varies considerably in different breast cancer subtypes (Honeth *et al.*, 2008). While the stem cell markers are enriched in basal like tumors, other tumor types, like $HER2^+$, shown only low levels of $CD44^+/CD24^-$ cells. In fact 32% of the tumors analyzed (n=232) did not have any $CD44^+/CD24^-$ cells. The authors conclude that potentially there are other stem cell markers that remain to be identified. The idea that other markers might exist is further supported by the fact that the $CD44^+/CD24^-$ profile is not correlated to tumorigenicity or distant metastases (Sheridan *et al.*, 2006; Fillmore and Kuperwasser, 2008). This indicates that while $CD44^+/CD24^-$ cells have tumor initiating capacities, the $CD44^+/CD24^-$ signature is not mandatory in cancer stem cells.

One surprising observation in the *in vivo* study was that the RPC32α knock-out cells were first detectable in mice, then seemed to disappear, before starting to grow into a tumor. This might indicate that from the originally injected 10 000 cells, only a small subset of cells were able to divide and multiply. The majority of cells seem to have died a few days into the experiment. Cells became visible again only after the rare surviving cells had started to multiply. Possibly the loss of RPC32α incapacitates cells to initiate a tumor. Only a few cells that had additional mutations were able to start a tumor. The fact that once the tumor started to grow, it was growing at approximately the same rate as the wild type tumor, would strengthen the hypothesis that RPC32α is primarily important for tumor initiation. However, the *in* vivo experiment has only be performed once. It will be necessary to repeat the experiment with different clones and observe if a similar decrease is visible before tumor onset begins.

Another indicator for RPC32α role in tumor initiation is the reduction of metastases by a factor 100. Metastases form when cells break away from a primary tumor, travel to distant locations in the body and initiate new tumors there. While tumor initiating capacity is a hallmark of cancer stem cells, the ability to break away from a growing tumor requires a transformation of the cell from attached to mobile. The epithelial–mesenchymal transition (EMT) allows the cell to detach from the surrounding cells and to move to a new site. Two well known markers for EMT are E-cadherin and vimentin, with

E-cadherin being a marker for the epithelial and vimentin a marker for the mesenchymal state (figure 18).

In the present study no alterations in the expression levels of either E-cadherin or vimentin could be detected using immunohistochemical staining between wild type and RPC32α knock-out tumors. The MDA-MB231 cell line is known to have an intrinsic mesenchymal phenotype, demonstrated by very low E-cadherin and strong vimentin levels. It is possible that the loss of RPC32α raised the levels of E-cadherin, but that overall levels remain too low for detection. Inversely the vimentin levels in the RPC32α KO cells could have been downregulated, but are still very high, thus no major change is detected. As mentioned before immunohistochemical staining might not be sensitive enough to detect subtle changes. Preliminary test using western blots have not led to conclusive results.

But EMT is regulated by a large number of factors and while E-cadherin and vimentin are certainly important markers, they are not definite. A study showed that by overexpressing the E-cadherin suppressor Snail, MDA-MB231 became more mobile, yet the levels of E-cadherin or vimentin did not change (Lundgren *et al.*, 2009). Inversely a knock-down of Snail reduced the migratory capacity of MDA-MB231 cells, and also of cell lines MCF7 and MDA-MB468, but again migration changed without significant alterations to E-cadherin or vimentin levels.

Another study reports that Snail is required for tumor growth and lymph node metastases in MDA-MB231 cells. Yet upon shRNA mediated knock-down of Snail, the authors only noted a modest increase of E-cadherin on an mRNA level and on a protein level E-cadherin remained undetectable (Olmeda *et al.*, 2007). Snail seems to be able to alter cell migration either independently of the EMT or using alternative regulators. These results strengthen the hypothesis that the tumors analyzed in the present study were altered by pathways independent of E-cadherin.

Besides Snail there are a number of transcription factors that regulate EMT. Among them other members of the Snail family, notably Slug and Twist, as well as the Zinc Finger E-Box Binding proteins 1 and 2 (ZEB1 and ZEB2) and E12/E47 (reviewed in Wang *et al.*, 2013 and Pattabiraman and Weinberg, 2014). It would be of great interest to test for the presence of those factors in the RPC32α knock-out cell lines to see if the loss of RPC32α has led to a reduction in any of those factors. It would also be important to analyze the metastases as well. For they originate from a cell that underwent either EMT or alternate transitions, in order to break away from the tumor bulk. Tissue samples have been recovered from different metastases for further analysis by immunohistochemistry or RT-qPCR.

Furthermore, metastases would have to be analyzed for stem cell markers such as $CD44^{+}/CD24^{-/low}$ or ALDH1, to find out if a loss of RPC32α has led to a loss of stem-cellness. But interpretation of the data would be difficult. As explained above the tumors formed in the RPC32α KO group potentially originated from cells that had additional mutations, able to overcome the constraints due a loss of

RPC32α. If it was indeed mutated cells that ultimately formed the tumor, than data from these cells does not accurately reflect the loss of RPC32α. It would reflect the effect of the random mutation that enables the cell to escape the RPC32α knock-down effect.

In order to answer the question whether RPC32α is involved in tumor growth, tumor initiation or both, a model would be needed, in which the endogenous RPC32α is replaced with an inducible version. The cells could be xenografted into mice, a tumor would form and then RPC32α would be switched off. If the tumor stops growing then RPC32α is important not only for tumor initiation, but also for tumor growth. However if the tumor keeps growing, but forms less metastases, then RPC32α is one of the key players in tumor initiation.

Undoubtedly more research is needed to uncover the role of RPC32α and the exact mechanisms by which it operates. Nonetheless, the results presented in this study are very encouraging and position RPC32α as an important factor in tumorigenesis and possibly as a key player in cancer stem cells. This together with the fact that RPC32α is overexpressed in triple-negative breast cancer, the most aggressive type of breast cancer, leaves hope that this research could benefit patients one day in form of a novel anti-cancer treatment.

5. Could RPC32α become a new drug target?

Currently no targeted therapy is available for triple negative cancers. Cytotoxic chemotherapy and surgery remain the only treatment options for patients with triple-negative breast cancer. For a while a promising candidate for targeted therapy seemed to be the monoclonal antibody bevacizumab. It binds the vascular endothelial growth factor (VEGF) and thereby inhibits angiogenesis. In phase III clinical trials bevacizumab was found to augment median progression free survival (PFS), but it did not show any effect on overall survival rates. This caused the US Food and Drug Administration (FDA) to revoke the agency's approval for bevacizumab in breast cancer (reviewd in Oualla *et al.*, 2017).

Recent studies to identify a possible treatment target involve immunotherapy, the epidermal growth factor receptor (EGFR), antiangiogenic agents, checkpoint kinase 1 inhibitors, the androgen receptor and factors of the DNA repair mechanisms (reviewed in Khosravi-Shahi *et al.*, 2017). Promising targets seem to be the poly-ADP-ribose-polymerases (PARPs). The PARPs are a family of 18 proteins that are involved in DNA repair (reviewed in Amé *et al.*, 2004). Since cancer cells divide rapidly, DNA replication is critical. By inhibiting PARP the base excision repair (BER), the single stranded break repair (SSBR) and the non-homologous end joining (NHEJ) pathways for DNA repair are blocked. This leaves cells with only homologous recombination (HR) to repair defective DNA. Given that many triple negative cancers harbor BRCA1/BRCA2 mutations, they are HR defective. Inhibiting PARPs leaves those cells without a functional DNA repair mechanism, which will ultimately result in cell

death. The advantage of PARP inhibitors is that they specifically target HR deficient cells, leaving healthy cells unscathed (reviewed in Livraghi and Garber, 2015).

One of the biggest challenges in the treatment of triple-negative cancer is its heterogeneous nature. There are numerous molecular subtypes of triple-negative cancer, which makes it very difficult to identify one single target that could act upon all these different cancer types. In this respect RPC32α seems to be a promising candidate. In the present study it was shown that POLR3G is overexpressed in triple-negative cancers from clinical breast cancer samples, making it a broad target for potentially all triple-negative breast cancers. Another challenge in the treatment of triple-negative breast cancers is their aggressiveness and metastases. This study shows that RPC32α is particularly effective in reducing metastases, possibly because loss of RPC32α most affects cancer stem cells.

Traditionally cancer therapies have targeted the primary tumor, without special treatment for cancer stem cells. These treatments will lead to a regression of the tumor as the bulk of the tumor cells die.

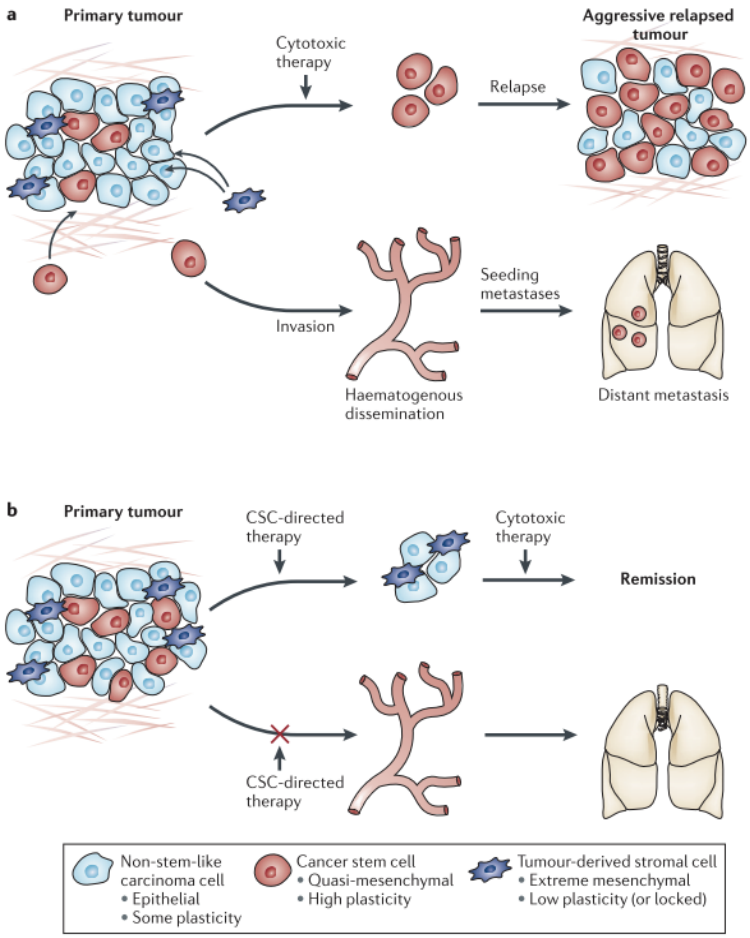

Figure 56: Impact of traditional versus cancer stem-cell directed therapy. (A) Traditional cytotoxic therapy targets most of the tumor cells, but cancer stem cells (red) can escape. Once the treatment stops a new tumor will form. Also it is cancer stem cells that lead to metastases, therefore cytotoxic therapy is not effective in preventing metastases. (B) If special cancer stem cell directed therapy is offered ahead of cytotoxic therapy, CSC are eliminated and tumors cannot be reinitiated after cytotoxic therapy stops. Furthermore CSC directed therapy would be able to prevent the cancer from spreading to other parts of the body.

Image from Pattabiraman and Weinberg, 2014

However, the surviving CSC are able to start a new tumor once the treatment stops, thus leading to tumor relapse (figure 56 A). Similarly CSC can start metastases, which are one of the main causes in cancer related death. If a drug would succeed to disable CSC, they could be used in combination with standard cytotoxic agents. First the CSC would eliminate any cell with tumor initiating capacity and in a second step the tumor cells are killed (figure 56 B). The present study found indications that RPC32α might be a good cancer stem cell target, however the *in vivo* experiment has been performed only once. Before the effect of RPC32α on tumorigenesis *in vivo* can be truly estimated, more mice experiments need to be done. It would be essential to repeat the experiment not only with another KO clone of the MDA-MB231 mother cell line, but also with clones of other triple-negative cell lines, such as BT-549 and MDA-MB468. The necessary clones have been created during the course of this study or are in the process of selection. Only if these different clones all lead to the same results *in vivo*, it can be estimated that RPC32α really might become a possible new drug target.

6. Outlook

The findings of this study are encouraging, but additional research needs to be done, to clearly identify the role of RPC32α. Most importantly the *in vivo* experiments need to be repeated with different KO clones. This is to confirm that the observed effects are neither clone nor lineage specific. The cell lines necessary for these controls have been created in the course of this study. But the RPC32α knock-out cell lines that have been created can serve a much wider purpose than just to repeat the *in vivo* experiments. In fact they are an indispensible tool to further elucidate the role of RPC32α in a tumor context. The analysis will have to be done at three levels: the cell lines, the tumors and the metastases.

On a cell line level, it would be of great interest to perform an RNA-seq analysis to identify alterations in the transcriptome of RPC32α KO cell lines. Using western blot or RT-qPCR, the levels of different stem cell markers or markers of EMT might be analyzed, to reveal possible changes in the RPC32α knock-out cell lines. Preliminary test on markers such as NANOG, OCT4 or E-cadherin have not produced any reliable results, but other testing conditions and other markers might help to reveal differences between the KO and the mother cell line. Furthermore, the RPC32α cell lines could be tested for stem cell markers *via* flow cytometry. This could show if the percentage of CSC in the KO cell lines has diminished compared to the mother cell line.

To test if the RPC32α KO cells have a reduced tumor initiating capacity, it would be of interest to perform a series of xenografts in mice with a decreasing number of injected cells. In the present study 10 000 cells were used to study tumor formation. Starting from 10 000 cells a series of five

xenografts with 5 000, 2500, 1000, 500 and 100 cells could be done, to identify the minimum number of cells necessary to start a tumor. If the number of cells necessary in the mother cell line is higher than in the KO cell line, than RPC32α KO cells indeed have a reduced tumor initiating ability.

Analyzes of the primary tumor and the metastases *via* immunohistochemistry and RT-qPCR will bring new insights into pathways altered by the loss of RPC32α and its potential effect on CSC. The necessary tissue extractions and conservation steps have been taken during this study. Markers that will be of particular interest include stem cell markers $CD44^{+}/CD24^{-/low}$ and ALDH1. Furthermore, markers of the EMT signaling pathway such as zona occludens 1 (ZO-1), N-cadherin or P-cadherin might show any transition from epithelial to mesenchymal phenotype or *vice versa*. The analysis of transcription factors like SNAIL, SLUG or TWIST will also help to elucidate possible connections between EMT and RPC32α.

Furthermore, rescue experiments should be performed with ectopically expressed RPC32α. RPC32β should be included in these rescue experiments to test if the observed effects are RPC32α specific or if at high enough doses RPC32β can rescue RPC32α KO cells. Also functional analysis of RPC32α has to include RPC32β, as maybe the reason why RPC32α promotes tumor growth is not in what RPC32α does, but rather in what it does not do. Possibly RPC32α does not act like an oncogene, but RPC32β acts like a tumor suppressor. In stem cells it would not be the down regulation of RPC32α that leads to differentiation, but rather the upregulation of RPC32β. Likewise in tumor cells it would be the loss of RPC32β that leads to tumor growth and not the presence of RPC32α. While the current data do not make this a likely scenario, it is a possibility that ought to be tested.

This study established breast cancer as a model for the characterization of RPC32α. But overexpression of POLR3G is not limited to breast cancer. According to the *gene enrichment profiler* databank of the Harvard Center for Computational and Integrative Biology, other cancers like melanoma, b cell lymphoma, colon and prostate cancer all have higher expression levels of POLR3G than breast cancers. It would be of interest to analyze the effect of an RPC32α KO in one of these cancers, particularly in melanomas. Possibly RPC32α could become a treatment target not only for triple-negative breast cancer, but also other cancers.

One of the objectives of this study was to establish a model to characterize RPC32α, this has been accomplished. Furthermore, the present study advanced the knowledge on RPC32α, notably by revealing its importance for tumorigenic growth *in vitro* and *in vivo*. During the course of this work, many important tools such as RPC32α KO cells have been created that will serve in future studies and thus help elucidate the role of RPC32α.

Chapter IV

Bibliography

Aasland, R. *et al.* (1996) 'The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional corepressor N-CoR and TFIIIB', *Trends in Biochemical Sciences*, 21(3), pp. 87–88. doi: 10.1016/S0968-0004(96)30009-1.

Abbas, S. *et al.* (2007) 'Serum 25-hydroxyvitamin D and risk of post-menopausal breast cancer-results of a large case-control study', *Carcinogenesis*, 29(1), pp. 93–99. doi: 10.1093/carcin /bgm240.

Abdelmohsen, K. *et al.* (2014) '7SL RNA represses p53 translation by competing with HuR', *Nucleic Acids Research*, 42(15), pp. 10099–10111. doi: 10.1093/nar/gku686.

Al-Hajj, M. *et al.* (2003) 'Prospective identification of tumorigenic breast cancer cells', *PNAS*, 100(7), pp. 3983–8. doi: 10.1073/pnas.0530291100.

Allison, D. S. and Hall, B. D. (1985) 'Effects of alterations in the 3' flanking sequence on in vivo and in vitro expression of the yeast SUP4-o tRNATyr gene', *The EMBO Journal*, 4(10), pp. 2657–64.

Amé, J.-C. *et al*. (2004) 'The PARP superfamily', *BioEssays*, 26(8), pp. 882–893. doi: 10.1002/bies.20085.

De Amicis, F. *et al.* (2010) 'Androgen receptor overexpression induces tamoxifen resistance in human breast cancer cells', *Breast Cancer Research and Treatment*, 121(1), pp. 1–11. doi: 10.1007/s10549-009-0436-8.

Andersson, M. G. *et al.* (2005) 'Suppression of RNA interference by adenovirus virus-associated RNA', *Journal of Virology*, 79(15), pp. 9556–65. doi: 10.1128/JVI.79.15.9556-9565.2005.

Arakawa, H. *et al.* (1995) 'Molecular cloning, characterization, and chromosomal mapping of a novel human gene (GTF3A) that is highly homologous to Xenopus transcription factor IIIA', *Cytogenetics and Cell Genetics*, 70(3–4), pp. 235–8.

Arimbasseri, A. G. *et al*. (2013) 'Transcription termination by the eukaryotic RNA polymerase III', *Biochimica et Biophysica Acta*, 1829(3–4), pp. 318–30. doi: 10.1016 /j.bbagrm.2012.10.006.

Arimbasseri, A. G. *et al*. (2014) 'Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation', *Transcription*, 5(1), p. e27369. doi: 10.4161/trns.27369.

Arrebola, R. *et al.* (1998) 'Tau91, an essential subunit of yeast transcription factor IIIC, cooperates with tau138 in DNA binding', *Molecular and Cellular Biology*, 18(1), pp. 1–9.

Badve, S. *et al.* (2010) 'Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists', *Modern Pathology*, 24, pp. 157–167. doi: 10.1038/modpathol.2010.200.

Bai, L. *et al.* (1996) 'Cloning and Characterization of the β Subunit of Human Proximal Sequence Element-Binding Transcription Factor and Its Involvement in Transcription of Small Nuclear RNA Genes by RNA Polymerases II and III', *Molecular and Cellular Biology*, 16(10), pp. 5419–5426.

Baker, R. E. *et al.* (1987) 'Gene size differentially affects the binding of yeast transcription factor X to two intragenic regions (tRNA transcription/RNA polymerase III/transcription factor IIIC)', *Biochemistry*, 84, pp. 8768–8772.

Balasubramanian, S. *et al.,* (2011) 'Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?', *Nature Reviews Drug Discovery*, 10. doi: 10.1038 /nrd3428.

Bane, A. *et al.* (2013) 'Clinical–pathologic significance of cancer stem cell marker expression in familial breast cancers', *Breast Cancer Research and Treatment*, 140(1), pp. 195–205. doi: 10.1007/s10549-013-2591-1.

Bao, S. *et al.* (2006) 'Glioma stem cells promote radioresistance by preferential activation of the DNA damage response', *Nature*, 444(7120), pp. 756–760. doi: 10.1038/nature05236.

Bardeleben, C. *et al*. (1994) 'Encounters of Saccharomyces cerevisiae RNA polymerase III with its transcription factors during RNA chain elongation', *Journal of Molecular Biology*, 235(4), pp. 1193–205.

Bark, C. *et al.* (1987) 'A distant enhancer element is required for polymerase III transcription of a U6 RNA gene', *Nature*, 328(6128), pp. 356–359. doi: 10.1038/328356a0.

Barski, A. *et al.* (2010) 'Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes', *Nature Structural & Molecular Biology*. NIH Public Access, 17(5), pp. 629–34. doi: 10.1038/nsmb.1806.

Bartholomew, B. *et al.* (1990) 'The subunit structure of Saccharomyces cerevisiae transcription factor IIIC probed with a novel photocrosslinking reagent', *The EMBO Journal*, 9(7), pp. 2197–205.

Bedrat, A. *et al*. (2016) 'Re-evaluation of G-quadruplex propensity with G4Hunter', *Nucleic Acids Research*, 44(4), p. gkw006. doi: 10.1093/nar/gkw006.

Behbod, F. *et al.* (2009) 'An intraductal human-in-mouse transplantation model mimics the subtypes of ductal carcinoma in situ', *Breast Cancer Research : BCR*, 11(5), p. R66. doi: 10.1186/bcr2358.

Bell, G. I. *et al.* (1977) 'Ribosomal RNA genes of Saccharomyces cerevisiae. I. Physical map of the repeating unit and location of the regions coding for 5 S, 5.8 S, 18 S, and 25 S ribosomal RNAs', *Journal of biological chemistry*, 252(22), pp. 8118–25.

Berger, W. *et al.* (2009) 'Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity', *Cellular and Molecular Life Sciences*, 66(1), pp. 43–61. doi: 10.1007/s00018-008-8364-z.

Bergsagel, D. E. and Valeriote, F. A. (1968) 'Growth Characteristics of a Mouse Plasma Cell Tumor', *Cancer Research*, 28(11), pp. 2187-96.

Betz, C. and Hall, M. N. (2013) 'Where is mTOR and what is it doing there?', *The Journal of Cell Biology*, 203(4).

Bieker, J. J. *et al*. (1985) 'Formation of a rate-limiting intermediate in 5S RNA gene transcription', *Cell*, 40(1), pp. 119–27.

Bogenhagen, D. F. *et al.* (1981) 'Nucleotide sequences in Xenopus 5S DNA required for transcription termination', *Cell*, 24(1), pp. 261–70. doi: 10.1016/0092-8674(81)90522-5.

Bogenhagen, D. F. *et al*. (1980) 'A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3′ border of the region', *Cell*, 19(1), pp. 27‑35. doi: 10.1016/0092-8674(80)90385-2.

Boissier, F. *et al.* (2015) 'Structural analysis of human RPC32β-RPC62 complex', *Journal of Structural Biology*, 192(3), pp. 313–319. doi: 10<.1016/j.jsb.2015.09.004.

Booy, E. P. *et al.* (2017) 'The long non-coding RNA BC200 (BCYRN1) is critical for cancer cell survival and proliferation', *Molecular Cancer*, 16(1), pp. 1–15. doi: 10.1186/s12943-017-0679-7.

Börner, T. *et al.* (2015) 'Chloroplast RNA polymerases: Role in chloroplast biogenesis', *Biochimica et Biophysica Acta - Bioenergetics*, 1847(9), pp. 761–769. doi: 10.1016 /J.BBABIO.2015.02.004.

Van Bortle, K. and Corces, V. G. (2012) 'tDNA insulators and the emerging role of TFIIIC in genome organization', *Transcription*, 3(6), pp. 277–284. doi: 10.4161/trns.21579.

Bozorgi, A. *et al.,* (2015) 'New Findings on Breast Cancer Stem Cells: A Review', *Journal of Breast Cancer*, 18(4), p. 303. doi: 10.4048/jbc.2015.18.4.303.

Braglia, P. *et al.,* (2005) 'Sequence context effects on oligo(dT) termination signal recognition by Saccharomyces cerevisiae RNA polymerase III', *The Journal of Biological Chemistry*, 280(20), pp. 19551–62. doi: 10.1074/jbc.M412238200.

Braun, B. R. *et al.* (1992) 'Topography of transcription factor complexes on the Saccharomyces cerevisiae 5 S RNA gene', *Journal of Molecular Biology*, 228(4), pp. 1063–77.

Brenton, J. D. *et al.* (2005) 'Molecular classification and molecular forecasting of breast cancer: ready for clinical application?', *Journal of Clinical Oncology*, 23(29), pp. 7350–60. doi: 10.1200 /JCO.2005.03.3845.

Bridier-Nahmias, A. *et al.* (2015) 'An RNA polymerase III subunit determines sites of retrotransposon integration', *Science*, 348(6234), pp. 585–588. doi: 10.1126/science.1259114.

Brogie, J. E. and Price, D. H. (2017) 'Reconstitution of a functional 7SK snRNP', *Nucleic Acids Research*, 45(11), pp. 1–17. doi: 10.1093/nar/gkx262.

Brow, D. A. and Guthrie, C. (1988) 'Spliceosomal RNA U6 is remarkably conserved from yeast to mammals', *Nature*, 334(6179), pp. 213–8. doi: 10.1038/334213a0.

Brow, D. A. and Guthrie, C. (1990) 'Transcription of a yeast U6 snRNA gene requires a polymerase III promoter element in a novel position', *Genes & Development*, 4(8), pp. 1345–56.

Brun, I. *et al*. (1997) 'Dual role of the C34 subunit of RNA polymerase III in transcription initiation', *The EMBO Journal*, 16(18), pp. 5730–41. doi: 10.1093/emboj /16.18.5730.

Buratowski, S. *et al.* (1992) 'A suppressor of TBP mutations encodes an RNA polymerase III transcription factor with homology to TFIIB', *Cell*, 71(2), pp. 221–30. doi: 10.1016/0092-8674(92)90351-C.

Burgess, R. R. *et al.* (1969) 'Factor Stimulating Transcription by RNA Polymerase', *Nature*, 221(5175), pp. 43–46. doi: 10.1038/221043a0.

Burgess, R. R. (1969) 'Separation and characterization of the subunits of ribonucleic acid polymerase', *The Journal of Biological Chemistry*, 244(22), pp. 6168–76.

Cailleau, R. *et al.* (1974) 'Breast Tumor Cell Lines From Pleural Effusions 1,2', *Journal of the National Cancer Institute*, 53(3), pp. 661–674.

Cailleau, R. *et al.* (1978) 'Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization', *In Vitro*, 14(11), pp. 911–15. doi: 10.1007/BF02616120.

Cairns, C. A. and White, R. J. (1998) 'p53 is a general repressor of RNA polymerase III transcription', The *EMBO Journal*, 17(11), pp. 3112–3123. doi: 10.1093/emboj/17.11.3112.

Camier, S. *et al*. (1995) 'The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes', *PNAS*, 92(20), pp. 9338–42.

Campbell, F. E. and Setzer, D. R. (1992) 'Transcription termination by RNA polymerase III: uncoupling of polymerase release from termination signal recognition', *Molecular and Cellular Biology*, 12(5), pp. 2260–72.

Canella, D. *et al*. (2010) 'Defining the RNA polymerase III transcriptome : Genome-wide localization of the RNA polymerase III transcription machinery in human cells', *Genome Research*, 20(6), pp. 710–721. doi: 10.1101/gr.101337.109.

Carbon, P. *et al.* (1987) 'A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II', *Cell*, 51(1), pp. 71–9.

Carbon, P. and Krol, A. (1991) 'Transcription of the Xenopus laevis selenocysteine tRNA(Ser)Sec gene: a system that combines an internal B box and upstream elements also found in U6 snRNA genes', *The EMBO Journal*, 10(3), pp. 599–606.

Cariati, M. *et al.* (2008) 'Alpha-6 integrin is necessary for the tumourigenicity of a stem cell-like subpopulation within the MCF7 breast cancer cell line', *International Journal of Cancer*, 122(2), pp. 298–304. doi: 10.1002/ijc.23103.

Carnevali, D. *et al.* (2017) 'Whole-genome expression analysis of mammalian-wide interspersed repeat elements in human cell lines', *DNA Research*, 24(1), pp. 59–69. doi: 10.1093/dnares/dsw048.

Carter, R. and Drouin, G. (2010) 'The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors', *Molecular Biology and Evolution*, 27(5), pp. 1035–43. doi: 10.1093/molbev/msp316.

Chakraborty, A. *et al.* (2012) 'Opening and closing of the bacterial RNA polymerase clamp', *Science*, 337(6094), pp. 591–5. doi: 10.1126/science.1218716.

Challice, J. M. and Segall, J. (1989) 'Transcription of the 5 S rRNA gene and Saccharomyces cerevisiae requires a promoter element at +1 and a 14-base pair internal control region', *Journal of Biological Chemistry*, 264(33), pp. 20060–20067.

Chang, D. D. and Clayton, D. A. (1987) 'A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication', The *EMBO journal*, 6(2), pp. 409–17.

Chaussivert, N. *et al.* (1995) 'Complex interactions between yeast TFIIIB and TFIIIC', *The Journal of Biological Chemistry*, 270(25), pp. 15353–8. doi: 10.1074/JBC.270.25.15353.

Chedin, S. *et al.* (1998) 'The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination', *Genes & Development*, 12(24), pp. 3857–3871. doi: 10.1101/gad.12.24.3857.

Chen, L.-L. and Yang, L. (2017) 'ALUternative Regulation for Gene Expression', *Trends in Cell Biology*, 27(7), pp. 480–490. doi: 10.1016/j.tcb.2017.01.002.

Chen, L. L. and Carmichael, G. G. (2008) 'Gene regulation by SINES and inosines: Biological consequences of A-to-I editing of Alu element inverted repeats', *Cell Cycle*, 7(21), pp. 3294–3301. doi: 10.4161/cc.7.21.6927.

Chen, W. *et al.* (1997) 'Expression of neural BC200 RNA in human tumours', *Journal of Pathology*, 183(3), pp. 345–51.

Chen, X. *et al.* (2003) 'The Ro autoantigen binds misfolded U2 small nuclear RNAs and assists mammalian cell survival after UV irradiation', *Current Biology*, 13(24), pp. 2206–11.

Chesnokov, I. *et al.* (1996) 'p53 inhibits RNA polymerase III-directed transcription in a promoter-dependent manner', *Molecular and Cellular Biology*, 16(12), pp. 7084–8.

Chia, K. *et al.* (2015) 'Targeting the Androgen Receptor in Breast Cancer', *Current Oncology Reports*, 17(2), p. 4. doi: 10.1007/s11912-014-0427-8.

Christov, C. P. *et al.* (2006) 'Functional requirement of noncoding Y RNAs for human chromosomal DNA replication', *Molecular and Cellular Biology*, 26(18), pp. 6993–7004.

Chu, W. M. *et al*. (1997) 'Palindromic sequences preceding the terminator increase polymerase III template activity', *Nucleic Acids Research*, 25(11), pp. 2077–82.

Ciliberto, G. *et al.* (1983) 'Common and interchangeable elements in the promoters of genes transcribed by RNA polymerase III', *Cell*, 32(3), pp. 725–733. doi: 10.1016/0092-8674(83)90058-2.

Clarke, E. M. *et al.* (1996) 'Regulation of the RNA polymerase I and III transcription systems in response to growth conditions', *The Journal of Biological Chemistry*, 271(36), pp. 22189–95. doi: 10.1074/JBC.271.36.22189.

Cogoi, S. *et al.* (2008) 'Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins', *Nucleic Acids Research*, 36(11), pp. 3765–80. doi: 10.1093/nar/gkn120.

Cogoi, S. and Xodo, L. E. (2006) 'G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription', *Nucleic Acids Research*, 34(9), pp. 2536–49. doi: 10.1093/nar/gkl286.

Colbert, T. and Hahn, S. (1992) 'A yeast TFIIB-related factor involved in RNA polymerase III transcription', *Genes & Development*, 6(10), pp. 1940–9.

Cole, C. *et al.* (2009) 'Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs', *RNA*, 15(12), pp. 2147–2160. doi: 10.1261/rna.1738409.

Coleman, M. P. *et al.* (2008) 'Cancer survival in five continents: a worldwide population-based study (CONCORD)', *The Lancet Oncology, 9(8), pp. 730-56*. doi: 10.1016/S1470.

Conti, A. *et al.* (2015) 'Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data', *Nucleic Acids Research*, 43(2), pp. 817–835. doi: 10.1093 /nar/gku1361.

Cook, J. A. *et al.* (2004) 'Oxidative stress, redox, and the tumor microenvironment', *Seminars in Radiation Oncology*, 14(3), pp. 259–266. doi: 10.1016/j.semradonc.2004.04.001.

Cooper, G. M. (2000) *The Cell: A Molecular Approach*. 2nd edition. Sunderland (MA): Sinauer Associates. ISBN-10: 0-87893-106-6

Cozzarelli, N. R. *et al.* (1983) 'Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes', *Cell*, 34(3), pp. 829–35.

Cramer, P. *et al.* (2000) 'Architecture of RNA Polymerase II and Implications for the Transcription Mechanism', *Science*, 288(5466), pp. 640–49.

Cramer, P. *et al.* (2001) 'Structural Basis of Transcription : RNA Polymerase II at 2.8 Ångstrom Resolution', *Science*, 292(5523), pp. 1863–1877. doi: 10.1126/science.1059493.

Cramer, P. *et al.* (2008) 'Structure of Eukaryotic RNA Polymerases', *Annual Review of Biophysics*, 37(1), pp. 337–352. doi: 10.1146/annurev.biophys.37.032807.130008.

Creighton, C. J. *et al.* (2009) 'Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features', *PNAS*, 106(33), pp. 13820–25.

Crighton, D. *et al.* (2003) 'p53 represses RNA polymerase III transcription by targeting TBP and inhibiting promoter occupancy by TFIIIB', *The EMBO Journal*, 22(11), pp. 2810–20. doi: 10.1093/emboj/cdg265.

Cuenca-López, M. D. *et al.* (2014) 'Phospho-kinase profile of triple negative breast cancer and androgen receptor signaling', *BMC Cancer*, 14(1), p. 302. doi: 10.1186/1471-2407-14-302.

Cufi, S. *et al.* (2012) 'Metformin-induced preferential killing of breast cancer initiating CD44+CD24-/low cells is sufficient to overcome primary resistance to trastuzumab in HER2+ human breast cancer xenografts', *Oncotarget*, 3(4), pp. 395–8. doi: 10.18632/oncotarget.488.

Dai, X. *et al.* (2015) 'Breast cancer intrinsic subtype classification, clinical use and future trends', *American Journal of Cancer Research*, 5(10), pp. 2929–43.

Danaei, G. *et al.* (2005) 'Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors', *The Lancet*, 366(9499), pp. 1784–1793. doi: 10.1016/S0140-6736(05)67725-2.

Das, G. *et al.* (1988) 'Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III', *The EMBO Journal*, 7(2), pp. 503–12.

Daskal, Y. *et al*. (1980) 'Isolation and partial characterization of perichromatin granules. A unique class of nuclear RNP particles', *Experimental Cell Research*, 126(1), pp. 39–46.

Daskalos, A. *et al.* (2009) 'Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer', *International Journal of Cancer*, 124(1), pp. 81–87. doi: 10.1002/ijc.23849.

Dean, N. and Berk, A. J. (1987) 'Separation of TFIIIC into two functional components by sequence specific DNA affinity chromatography', *Nucleic Acids Research*, 15(23), pp. 9895–907.

Deininger, P. *et al.* (2011) 'Alu elements: know the SINEs', *Genome Biology*, 12(12), p. 236. doi: 10.1186/gb-2011-12-12-236.

Dent, R. *et al.* (2007) 'Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence', *Clinical Cancer Research*, 13(15).

Deprez, E. *et al.* (1999) 'A subunit of yeast TFIIIC participates in the recruitment of TATA-binding protein', *Molecular and Cellular Biology*, 19(12), pp. 8042–51.

Dhahbi, J. M. *et al.* (2014) 'Deep Sequencing of Serum Small RNAs Identifies Patterns of 5′ tRNA Half and YRNA Fragment Expression Associated with Breast Cancer', *Biomarkers in cancer*, 6, pp. 37–47. doi: 10.4137/BiC.s20764.

Dieci, G. *et al.* (2007) 'The expanding RNA polymerase III transcriptome', *Trends in Genetics*, 23(12), pp. 614–22. doi: 10.1016/j.tig.2007.09.001.

Dieci, G. *et al.* (2013) 'Identification of RNA polymerase III-transcribed genes in eukaryotic genomes', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(3–4), pp. 296–305. doi: 10.1016/j.bbagrm.2012.09.010.

Dieci, G. and Sentenac, A. (1996) 'Facilitated recycling pathway for RNA polymerase III', *Cell*, 84(2), pp. 245–52.

Domitrovich, A. M. and Kunkel, G. R. (2003) 'Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies', *Nucleic Acids Research*, 31(9), pp. 2344–2352. doi: 10.1093/nar/gkg331.

Donze, D. (2012) 'Extra-transcriptional functions of RNA Polymerase III complexes: TFIIIC as a potential global chromatin bookmark', *Gene*, 493(2), pp. 169–175. doi: 10.1016/j.gene.2011.09.018.

Ducrot, C. *et al.* (2006) 'Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors', *The Journal of Biological Chemistry*, 281(17), pp. 11685–92. doi: 10.1074/jbc.M600101200.

Dumay-Odelot, H. *et al.* (2002) 'Multiple roles of the tau131 subunit of yeast transcription factor IIIC (TFIIIC) in TFIIIB assembly', *Molecular and Cellular Biology*, 22(1), pp. 298–308.

Dumay-Odelot, H. *et al.* (2007) 'Identification, molecular cloning, and characterization of the sixth subunit of human transcription factor TFIIIC', *The Journal of Biological Chemistry*, 282(23), pp. 17179–89. doi: 10.1074/jbc.M611542200.

Dumay, H. *et al.* (1999) 'Interaction between yeast RNA polymerase III and transcription factor TFIIIC via ABC10alpha and tau131 subunits', *The Journal of Biological Chemistry*, 274(47), pp. 33462–8.

Durrieu-Gaillard, S. *et al.* (in preparation) 'Regulation of RNA polymerase III transcription during transformation of human IMR90 fibroblasts with defined genetic elements', *Cell Cycle*.

Eddy, J. *et al.* (2011) 'G4 motifs correlate with promoter-proximal transcriptional pausing in human genes', *Nucleic Acids Research*, 39(12), pp. 4975–4983. doi: 10.1093/nar/gkr079.

Eddy, J. and Maizels, N. (2008) 'Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes', *Nucleic Acids Research*, 36(4), pp. 1321–1333. doi: 10.1093/nar/gkm1138.

Eilers, M. and Eisenman, R. N. (2008) 'Myc's broad reach', *Genes & Development*, 22(20), pp. 2755–2766. doi: 10.1101/gad.1712408.

Elayat, G. *et al*. (2010) 'Cell turnover in apocrine metaplasia and apocrine adenosis of the breast', *Annals of Diagnostic Pathology*, 14(1), pp. 1–7. doi: 10.1016/j.anndiagpath.2009.05.001.

Elbarbary, R. A. *et al*. (2016) 'Retrotransposons as regulators of gene expression', *Science*, 351(6274), p. aac7247. doi: 10.1126/science.aac7247.

Engel, C. *et al.* (2013) 'RNA polymerase I structure and transcription regulation', *Nature*, 502(7473), pp. 650–5. doi: 10.1038/nature12712.

Engelke, D. R. *et al.* (1980) 'Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes', *Cell*, 19(3), pp. 717–28.

Englert, M. *et al.* (2004) 'Novel upstream and intragenic control elements for the RNA polymerase III-dependent transcription of human 7SL RNA genes', *Biochimie*, 86(12), pp. 867–74. doi: 10.1016/j.biochi.2004.10.012.

Enver, T. *et al.* (2005) 'Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells', *Human Molecular Genetics*, 14(21), pp. 3129–40. doi: 10.1093/hmg/ddi345.

Ezkurdia, I. *et al.* (2014) 'Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes', *Human Molecular Genetics*, 23(22), pp. 5866–78. doi: 10.1093/hmg/ddu309.

Farmer, P. *et al.* (2005) 'Identification of molecular apocrine breast tumours by microarray analysis', *2Oncogene*, 24(29), pp. 4660–71. doi: 10.1038/sj.onc.1208561.

Felton-Edkins, Z. A. and White, R. J. (2002) 'Multiple mechanisms contribute to the activation of RNA polymerase III transcription in cells transformed by papovaviruses', *Journal of Biological Chemistry*, 277(50), pp. 48182–48191. doi: 10.1074/jbc.M201333200.

Ferlay, J. *et al.* (no date) *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*, *Internet*. Available at: http://globocan.iarc.fr (Accessed: 15 August 2017).

Fernández-Tornero, C. *et al.* (2007) 'Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III', *Molecular Cell*, 25(6), pp. 813–23. doi: 10.1016/j.molcel.2007.02.016.

Fernández-Tornero, C. *et al.* (2010) 'Conformational flexibility of RNA polymerase III during transcriptional elongation', *The EMBO Journal*, 29(22), pp. 3762–3772. doi: 10.1038/emboj.2010.266.

Fernández-Tornero, C. *et al.* (2013) 'Crystal structure of the 14-subunit RNA polymerase I', *Nature*, 502(7473), pp. 644–649. doi: 10.1038/nature12636.

Ferrari, R. *et al.* (2004) 'Distinct roles of transcription factors TFIIIB and TFIIIC in RNA polymerase III transcription reinitiation', *PNAS*, 101(37), pp. 13442–7. doi: 10.1073/pnas.0403851101.

Ferri, M. L. *et al.* (2000) 'A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIIB70', *Molecular and Cellular Biology*, 20(2), pp. 488–95.

Fillmore, C. M. and Kuperwasser, C. (2008) 'Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy', *Breast Cancer Research : BCR*, 10(2), p. R25. doi: 10.1186/bcr1982.

Fleming, J. M. *et al.* (2010) 'Local regulation of human breast xenograft models', *Journal of cellular physiology*. NIH Public Access, 224(3), pp. 795–806. doi: 10.1002/jcp.22190.

Fok, V. *et al*. (2006) 'Epstein-Barr virus noncoding RNAs are confined to the nucleus, whereas their partner, the human La protein, undergoes nucleocytoplasmic shuttling', *The Journal of cell biology*, 173(3), p. 319.

Ford, D. *et al.* (1998) 'Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium', *American journal of human genetics*, 62(3), pp. 676–89.

Foulkes, W. D *et al*. (2010) 'Triple-Negative Breast Cancer', *N Engl J Med*, 20363(11), pp. 1938–48.

Fuchs, H. *et al.* (2014) 'miR-27 Negatively Regulates Pluripotency-Associated Genes in Human Embryonal Carcinoma Cells', *PloS ONE*, 9(11), p. e111637. doi: 10.1371/journal.pone.0111637.

Gabrielsen, O. S. *et al.* (1989) 'Two polypeptide chains in yeast transcription factor tau interact with DNA', *The Journal of Biological Chemistry*, 264(13), pp. 7505–11.

Galli, G. *et al*. (1981) 'Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements', *Nature*, 294(5842), pp. 626–31.

Geslain, R. and Pan, T. (2010) 'Functional Analysis of Human tRNA Isodecoders', *Journal of Molecular Biology*, 396(3), pp. 821–831. doi: 10.1016/j.jmb.2009.12.018.

Ghosh, R. *et al.* (2011) 'Trastuzumab has preferential activity against breast cancers driven by HER2 homodimers', *Cancer Research*. NIH Public Access, 71(5), pp. 1871–82. doi: 10.1158/0008-5472.CAN-10-1872.

Gill, T. *et al.* (2004) 'RNase MRP cleaves the CLB2 mRNA to promote cell cycle progression: novel method of mRNA degradation', *Molecular and Cellular Biology*, 24(3), pp. 945–53.

Ginestier, C. *et al.* (2007) 'ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome', *Cell Stem Cell*, 1(5), pp. 555–567. doi: 10.1016/j.stem.2007.08.014.

Gingold, H. *et al.* (2014) 'A Dual Program for Translation Regulation in Cellular Proliferation and Differentiation', *Cell*, 158, pp. 1281–1292. doi: 10.1016/j.cell.2014.08.011.

Ginsberg, A.M. *et al*. (1984) 'Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development', *Cell*, 39(3 Pt 2), pp. 479–89.

Goffeau, A. *et al.* (1996) 'Life with 6000 genes', *Science*, 274(5287), pp. 563–7.

Gogolevskaya, I.K. and Kramerov, D. A. (2010) '4.5SI RNA genes and the role of their 5'-flanking sequences in the gene transcription', *Gene*, 451(1–2), pp. 32–7. doi: 10.1016/j.gene.2009.11.007.

Goldfarb, K. C. and Cech, T. R. (2017) 'Targeted CRISPR disruption reveals a role for RNase MRP RNA in human preribosomal RNA processing', *Genes & Development*, 31(1), pp. 59–71. doi: 10.1101/gad.286963.116.

Gomez-Roman, N. *et al.* (2003) 'Direct activation of RNA polymerase III transcription by c-Myc', *Nature*, 421(6920), pp. 290–4. doi: 10.1038/nature01327.

Gong, B. *et al.* (2013) 'Caprin-1 is a novel microRNA-223 target for regulating the proliferation and invasion of human breast cancer cells', *Biomedicine & pharmacotherapy = Biomédecine & pharmacothérapie*, 67(7), pp. 629–36. doi: 10.1016/j.biopha.2013.06.006.

Gong, C. and Maquat, L. E. (2011) 'lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements', *Nature*, 470(7333), pp. 284–8. doi: 10.1038/nature09701.

Goodarzi, H. *et al.* (2015) 'Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement', *CELL*, 161, pp. 790–802. doi: 10.1016/j.cell.2015.02.053.

Goodarzi, H. *et al.* (2016) 'Modulated expression of specific tRNAs drives gene expression and cancer progression', *Cell*, 165(6), pp. 1416–1427. doi: 10.1016/j.cell.2016.05.046.

Goodenbour, J. M. and Pan, T. (2006) 'Diversity of tRNA genes in eukaryotes', *Nucleic Acids Research*, 34(21), pp. 6137–6146. doi: 10.1093/nar/gkl725.

Goodfellow, S. J. *et al.* (2006) 'Regulation of RNA polymerase III transcription during hypertrophic growth', *The EMBO Journal*, 25(7), pp. 1522–33. doi: 10.1038/sj.emboj.7601040.

Goodfellow, S. J. and Zomerdijk, J. C. B. M. (2012) 'Basic Mechanisms in RNA Polymerase I Transcription of the Ribosomal RNA Genes'. doi: 10.1007/978-94-007-4525-4_10.

Goodier, J. L. and Maraia, R. J. (1998) 'Terminator-specific recycling of a B1-Alu transcription complex by RNA polymerase III is mediated by the RNA terminus-binding protein La', *The Journal of Biological Chemistry*, 273(40), pp. 26110–6.

Gottesfeld, J. M. *et al*. (1996) 'Transcriptional activation of RNA polymerase III-dependent genes by the human T-cell leukemia virus type 1 tax protein', *Molecular and Cellular Biology*, 16(4), pp. 1777–85.

Gottlieb, E. and Steitz, J. A. (1989) 'The RNA binding protein La influences both the accuracy and the efficiency of RNA polymerase III transcription in vitro', *The EMBO Journal*, 8(3), pp. 841–50.

Gouge, J. *et al.* (2017) 'Molecular mechanisms of Bdp1 in TFIIIB assembly and RNA polymerase III transcription initiation', *Nature Communications*, 8(1), p. 130. doi: 10.1038/s41467-017-00126-1.

Grand, C. L. *et al.* (2002) 'The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo 1', *Molecular Cancer Therapeutics*, 1(8), pp. 565–73.

Greenleaf, A. L. *et al*. (1986) 'Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome', *PNAS*, 83(10), pp. 3391–4.

Grewal, S. S. (2014) 'Why should cancer biologists care about tRNAs? TRNA synthesis, mRNA translation and the control of growth', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1849(7), pp. 898–907. doi: 10.1016/j.bbagrm.2014.12.005.

Gunnery, S. *et al.* (1999) 'Termination sequence requirements vary among genes transcribed by RNA polymerase III 1 1Edited by J. Karn', *Journal of Molecular Biology*, 286(3), pp. 745–757. doi: 10.1006/jmbi.1998.2518.

Guo, J. and Price, D. H. (2013) 'RNA polymerase II transcription elongation contro', *Chemical reviews*, 113(11), pp. 8583–603.

Gupta, P. B. *et al.* (2011) 'Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells', *Cell*, 146(4), pp. 633–644. doi: 10.1016/j.cell.2011.07.026.

Haeusler, R. A. and Engelke, D. R. (2006) 'Spatial organization of transcription by RNA polymerase III', *Nucleic Acids Research*, 34(17), pp. 4826–36.

Hahn, S. *et al.* (1989) 'Yeast TATA-Binding Protein TFIID Binds to TATA Elements with Both Consensus and Nonconsensus DNA Sequences', *PNAS*, pp. 5718–5722. doi: 10.2307/34191.

Hall, A. E. *et al*. (2013) 'Y RNAs: Recent developments', *Biomolecular Concepts*, 4(2), pp. 103–110. doi: 10.1515/bmc-2012-0050.

Hall, J. M. *et al.* (1990) 'Linkage of early-onset familial breast cancer to chromosome 17q21', *Science*, 250(4988), pp. 1684–9.

Hamada, M. *et al.* (2000) 'Transcription termination by RNA polymerase III in fission yeast. A genetic and biochemically tractable model system', *The Journal of Biological Chemistry*, 275(37), pp. 29076–81. doi: 10.1074/jbc.M003980200.

Harismendy, O. *et al.* (2003) 'Genome-wide location of yeast RNA polymerase III transcription machinery', *The EMBO Journal*, 22(18), pp. 4738–47. doi: 10.1093/emboj/cdg466.

Haurie, V. *et al.* (2010) 'Two isoforms of human RNA polymerase III with specific functions in cell growth and transformation', *PNAS*, 107(9), pp. 4176–81. doi: 10.1073/pnas.0914980107.

Helbo, A. S. *et al.* (2017) 'Nucleosome Positioning and NDR Structure at RNA Polymerase III Promoters', *Scientific Reports*, p. 41947. doi: 10.1038/srep41947.

Hennessy, B. T. *et al.* (2009) 'Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics', *Cancer Research*, 69(10), pp. 4116–4124. doi: 10.1158/0008-5472.CAN-08-3441.

Henry, R. W. *et al.* (1995) 'A TBP–TAF complex required for transcription of human snRNA genes by RNA polymerases II and III', *Nature*, 374(6523), pp. 653–656. doi: 10.1038/374653a0.

Henry, R. W. *et al.* (1996) 'Cloning and characterization of SNAP50, a subunit of the snRNA-activating protein complex SNAPc', *The EMBO Journal*, 15(24), pp. 7129–36.

Henry, R. W. *et al.* (1998) 'SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III', *Genes & Development*, 12(17), p. 2664.

Heppner, G. H. *et al.* (1983) 'Mammary tumor heterogeneity', *Symposium on Fundamental Cancer Research*, 36, pp. 209–21.

Hirsch, H. A. *et al.* (2004) 'Distinct Mechanisms for Repression of RNA Polymerase III Transcription by the Retinoblastoma Tumor Suppressor Protein', *Molecular and Cellular Biology*, 24(13), pp. 5989–5999. doi: 10.1128/MCB.24.13.5989-5999.2004.

Hirsch, H. A. *et al.* (2009) 'Metformin Selectively Targets Cancer Stem Cells, and Acts Together with Chemotherapy to Block Tumor Growth and Prolong Remission', *Cancer Research*, 69(19), pp. 7507–7511. doi: 10.1158/0008-5472.CAN-09-2994.

Hoffmann, N. A. *et al.* (2015) 'Molecular structures of unbound and transcribing RNA polymerase III', *Nature*, 528(7581), pp. 231–236. doi: 10.1038/nature16143.

Hoffmann, N. A. *et al.* (2016) 'Transcribing RNA polymerase III observed by electron cryomicroscopy', *FEBS Journal*, 283, pp. 2811–2819. doi: 10.1111/febs.13732.

Hofstetter, H. *et al*. (1981) 'A split promoter for a eucaryotic tRNA gene', *Cell*, 24(2), pp. 573–585. doi: 10.1016/0092-8674(81)90348-2.

Holliday, D. L. and Speirs, V. (2011) 'Choosing the right cell line for breast cancer research', *Breast Cancer Research*, 215(13), pp. 1–7. doi: 10.1186/bcr2889.

Honeth, G. *et al.* (2008) 'The CD44+/CD24-phenotype is enriched in basal-like breast tumors', *Breast Cancer Research*, 10(3), p. R53. doi: 10.1186/bcr2108.

Horikoshi, M. *et al.* (1989) 'Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box', *Nature*, 341(6240), pp. 299–303. doi: 10.1038/341299a0.

Horiuchi, D. *et al.* (2012) 'MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition', *The Journal of experimental medicine*, 209(4), pp. 679–96. doi: 10.1084/jem.20111512.

Hsieh, Y. J. *et al.* (1999) 'Cloning and characterization of two evolutionarily conserved subunits (TFIIIC102 and TFIIIC63) of human TFIIIC and their involvement in functional interactions with TFIIIB and RNA polymerase III', *Molecular and Cellular Biology*, 19(7), pp. 4944–52.

Hsieh, Y. J., *et al.* (1999) 'The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity', *Molecular and Cellular Biology*, 19(11), pp. 7697–704.

Hu, H.-L. *et al.* (2015) 'A Region of Bdp1 Necessary for Transcription Initiation That Is Located within the RNA Polymerase III Active Site Cleft', *Molecular and Cellular Biology*, 35(16), pp. 2831–40. doi: 10.1128/MCB.00263-15.

Hu, P. *et al*. (2003) 'A minimal RNA polymerase III transcription system from human cells reveals positive and negative regulatory roles for CK2', *Molecular Cell*, 12(3), pp. 699–709.

Hu, S. *et al.* (2012) 'Signals from noncoding RNAs: Unconventional roles for conventional pol III transcripts', *International Journal of Biochemistry and Cell Biology*, 44(11), pp. 1847–1851. doi: 10.1016/j.biocel.2012.07.013.

Hu, Z. *et al.* (2006) 'The molecular portraits of breast tumors are conserved across microarray platforms', *BMC genomics*, 7(96). doi: 10.1186/1471-2164-7-96.

Huet, J. and Sentenac, A. (1992) 'The TATA-binding protein participates in TFIIIB assembly on tRNA genes', *Nucleic Acids Research*, 20(24), pp. 6451–4.

Iben, J. R. *et al.* (2011) 'Point mutations in the Rpb9-homologous domain of Rpc11 that impair transcription termination by RNA polymerase III', *Nucleic Acids Research*, 39(14), pp. 6100–6113. doi: 10.1093/nar/gkr182.

Inic, Z. *et al.* (2014) 'Difference between Luminal A and Luminal B Subtypes According to Ki-67, Tumor Size, and Progesterone Receptor Negativity Providing Prognostic Information', *Clinical Medicine Insights: Oncology*, 8, pp. 107–111. doi: 10.4137/CMO.s18006.

Jang, S. *et al.* (2017) 'Regulation of BC200 RNA-mediated translation inhibition by hnRNP E1 and E2', *FEBS Letters*, 591(2), pp. 393–405. doi: 10.1002/1873-3468.12544.

Jasiak, A. J. *et al.* (2006) 'Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model', *Molecular Cell*, 23(1), pp. 71–81. doi: 10.1016/j.molcel.2006.05.013.

Jawdekar, G. W. and Henry, R. W. (2008) 'Transcriptional regulation of human small nuclear RNA genes', *Biochimica et biophysica acta*, 1779(5), pp. 295–305. doi: 10.1016/j.bbagrm.2008.04.001.

Jin, S. *et al.* (2016) 'A Novel role for miR-1305 in Regulation of Pluripotency-Differentiation Balance, Cell Cycle and Apoptosis in Human Pluripotent Stem Cells', *Stem Cells*, pp. 1–15. doi: 10.1002/stem.2444.

Joazeiro, C. A. *et al*. (1994) 'Identical components of yeast transcription factor IIIB are required and sufficient for transcription of TATA box-containing and TATA-less genes', *Molecular and Cellular Biology*, 14(4), pp. 2798–808.

Johnson, S.A.S. *et al*. (2008) 'Enhanced RNA polymerase III-dependent transcription is required for oncogenic transformation', *The Journal of Biological Chemistry*, 283(28), pp. 19184–91. doi: 10.1074/jbc.M802872200.

Kantidakis, T. *et al.* (2010) 'mTOR associates with TFIIIC, is found at tRNA and 5S rRNA genes, and targets their repressor Maf1', *PNAS*, 107(26), pp. 11823–8. doi: 10.1073/pnas.1005188107.

Kapitonov, V. V and Jurka, J. (2003) 'A novel class of SINE elements derived from 5S rRNA', *Molecular biology and evolution*, 20(5), pp. 694–702. doi: 10.1093/molbev/msg075.

Kassavetis, G. A. *et al.* (1990) 'S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors', *Cell*, 60(2), pp. 235–45.

Kassavetis, G. A. *et al.* (1991) 'Two essential components of the Saccharomyces cerevisiae transcription factor TFIIIB: transcription and DNA-binding properties', *PNAS*, 88(16), pp. 7308–12. doi: 10.1073/PNAS.88.16.7308.

Kassavetis, G. A. *et al.* (1992) 'The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB', *Cell*, 71(6), pp. 1055–64.

Kassavetis, G. A. *et al.* (1995) 'Cloning, expression, and function of TFC5, the gene encoding the B" component of the Saccharomyces cerevisiae RNA polymerase III transcription factor TFIIIB', *PNAS*, 92(21), pp. 9786–90.

Kassavetis, G. A. *et al.* (2003) 'The Role of Transcription Initiation Factor IIIB Subunits in Promoter Opening Probed by Photochemical Cross-linking', *Journal of Biological Chemistry*, 278(20), pp. 17912–17917. doi: 10.1074/jbc.M300743200.

Kassavetis, G. A. *et al*. (2006) 'Mapping the Principal Interaction Site of the Brf1 and Bdp1 Subunits of Saccharomyces cerevisiae TFIIIB', *Journal of Biological Chemistry*, 281(20), pp. 14321–14329. doi: 10.1074/jbc.M601702200.

Kassavetis, G. A. *et al*. (2010) 'The C53/C37 Subcomplex of RNA Polymerase III Lies Near the Active Site and Participates in Promoter Opening', *Journal of Biological Chemistry*, 285(4), pp. 2695–2706. doi: 10.1074/jbc.M109.074013.

Kedersha, N. L. and Rome, L. H. (1986) 'Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA', *Journal of cell biology*, 103(3), pp. 699–709.

Kenneth, N. S. *et al.* (2007) 'TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription', *PNAS*, 104(38), pp. 14917–22. doi: 10.1073/pnas.0702909104.

Kenneth, N. S. *et al*. (2008) 'Recruitment of RNA polymerase III in vivo', *Nucleic Acids Research*, 36(11), pp. 3757–64. doi: 10.1093/nar/gkn272.

Khattar, E. *et al.* (2016) 'Telomerase reverse transcriptase promotes cancer cell proliferation by augmenting tRNA expression', 126(17), pp. 1–16. doi: 10.1172/JCI86042DS1.

Khatter, H., *et al*. (2017) 'RNA polymerase I and III: similar yet unique', *Current Opinion in Structural Biology*, 47, pp. 88–94. doi: 10.1016/j.sbi.2017.05.008.

Kheir, E. G. A. and Krude, T. (2017) 'Non-coding Y RNAs associate with early replicating euchromatin concordantly with the origin recognition complex (ORC)', *Journal of Cell Science*, 45, p. jcs.197566. doi: 10.1242/jcs.197566.

Khoo, B. *et al*. (1994) 'Conserved functional domains of the RNA polymerase III general transcription factor BRF', *Genes & Development*, 8(23), pp. 2879–90.

Khosravi-Shahi, P., *et al*. (2017) 'Metastatic triple negative breast cancer: Optimizing treatment options, new and emerging targeted therapies', *Asia-Pacific Journal of Clinical Oncology*, e_pub ahea. doi: 10.1111/ajco.12748.

Kikovska*, E. et al.* (2007) 'Eukaryotic RNase P RNA mediates cleavage in the absence of protein', *PNAS*, 104(7), pp. 2062–7.

Kim, Y. *et al.* (1993) 'Crystal structure of a yeast TBP/TATA-box complex', *Nature*, 365(6446), pp. 512–520. doi: 10.1038/365512a0.

Kise, K. *et al.* (2016) 'Tumor microenvironment for cancer stem cells', *Advanced Drug Delivery Reviews*, 99(Pt B), pp. 197–205. doi: 10.1016/j.addr.2015.08.005.

Kitajewski, J. *et al.* (1986) 'Adenovirus VAI RNA antagonizes the antiviral action of interferon by preventing activation of the interferon-induced eIF-2α kinase', *Cell*, 45(2), pp. 195–200. doi: 10.1016/0092-8674(86)90383-1.

Kondrashov, A. V. *et al.* (2005) 'Inhibitory effect of naked neural BC1 RNA or BC200 RNA on eukaryotic in vitro translation', *Journal of Molecular Biology*, 353(1), pp. 88–103.

Kowalski, M. P. and Krude, T. (2015) 'Functional roles of non-coding Y RNAs', *International Journal of Biochemistry and Cell Biology*, 66, pp. 20–29. doi: 10.1016/j.biocel.2015.07.003.

Kramerov, D. A. and Vassetzky, N. S. (2011) 'Origin and evolution of SINEs in eukaryotic genomes', *Heredity*, 107(6), pp. 487–95. doi: 10.1038/hdy.2011.43.

Kriegs, J. O. *et al.* (2007) 'Evolutionary history of 7SL RNA-derived SINEs in Supraprimates', *Trends in Genetics*, 23(4), pp. 158–161. doi: 10.1016/j.tig.2007.02.002.

Krüger, W. and Benecke, B. J. (1987) 'Structural and functional analysis of a human 7 S K RNA gene', *Journal of Molecular Biology*, 195(1), pp. 31–41. doi: 10.1016/0022-2836(87)90325-1.

Kuhn, C.-D. *et al.* (2007) 'Functional Architecture of RNA Polymerase I', *Cell*, 131(7), pp. 1260–1272. doi: 10.1016/j.cell.2007.10.051.

Kundu, T. K. *et al*. (1999) 'Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity', *Molecular and Cellular Biology*, 19(2), pp. 1605–15.

Kunkel, G. R. *et al.* (1986) 'U6 small nuclear RNA is transcribed by RNA polymerase III', *PNAS*, 83(22), pp. 8575–79.

Kunkel, G. R. and Pederson, T. (1988) 'Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used', *Genes & Development*, 2(2), pp. 196–204.

L'Etoile, N. D. *et al.* (1994) 'Human transcription factor IIIC box B binding subunit', *PNAS*, 91(5), pp. 1652–6.

Laferté, A. *et al.* (2006) 'The transcriptional activity of RNA polymerase I is a key determinant for the leve of all ribosome components', *Genes & Development*, 20, pp. 2030–2040. doi: 10.1101/gad.386106.

Lagna, G. *et al.* (1994) 'Cloning and characterization of an evolutionarily divergent DNA-binding subunit of mammalian TFIIIC', *Molecular and Cellular Biology*, 14(5), pp. 3053–64.

Lalloo, F. and Evans, D. G. (2012) 'Familial Breast Cancer', *Clinical Genetics*, 82(2), pp. 105–114. doi: 10.1111/j.1399-0004.2012.01859.x.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Landrieux, E. *et al.* (2006) 'A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation', *The EMBO Journal*, 25(1), pp. 118–128. doi: 10.1038/sj.emboj.7600915.

Lapidot, T. *et al.* (1994) 'A cell initiating human acute myeloid leukaemia after transplantation into SCID mice', *Nature*, 367(6464), pp. 645–648. doi: 10.1038/367645a0.

Larminie, C. G. *et al.* (1999) 'Activation of RNA polymerase III transcription in cells transformed by simian virus 40', *Molecular and Cellular Biology*, 19(7), pp. 4927–34.

Lasfargues, E. Y. *et al*. (1978) 'Isolation of two human tumor epithelial cell lines from solid breast carcinomas', *Journal of the National Cancer Institute*, 61(4), pp. 967–978. doi: 10.1093/jnci/61.4.967.

Layat, E., *et al*. (2013) 'Structure, function and regulation of Transcription Factor IIIA: From Xenopus to Arabidopsis', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(3–4), pp. 274–282. doi: 10.1016/j.bbagrm.2012.10.013.

Lee, K. *et al.* (2011) 'Precursor miR-886, a novel noncoding RNA repressed in cancer, associates with PKR and modulates its activity', *RNA*, 17(6), pp. 1076–89. doi: 10.1261/rna.2701111.

Lee, N. *et al*. (2012) 'AUF1/hnRNP D is a novel protein partner of the EBER1 noncoding RNA of Epstein-Barr virus', *RNA*, 18(11), pp. 2073–82.

Lee, Y. S. (2015) 'A Novel Type of Non-coding RNA, nc886, Implicated in Tumor Sensing and Suppression', *Genomics & informatics*, 13(2), pp. 26–30. doi: 10.5808/GI.2015.13.2.26.

Lefebvre, O. *et al.* (1992) 'TFC3: gene encoding the B-block binding subunit of the yeast transcription factor IIIC', *PNAS*, 89(21), pp. 10512–6.

Lefèvre, S. *et al.* (2011) 'Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation', *Nature Structural & Molecular Biology*, 18(3), pp. 352–8. doi: 10.1038/nsmb.1996.

Lehmann-Che, J. *et al.* (2013) 'Molecular apocrine breast cancers are aggressive estrogen receptor negative tumors overexpressing either HER2 or GCDFP15', *Breast Cancer Research*, 15(3), p. R37. doi: 10.1186/bcr3421.

Lehmann, B. D. *et al.* (2011) 'Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies', *The Journal of Clinical Investigation*, 121(7), pp. 2750–67. doi: 10.1172/JCI45014.

Lehmann, B. D. *et al.* (2014) 'PIK3CA mutations in androgen receptor-positive triple negative breast cancer confer sensitivity to the combination of PI3K and androgen receptor inhibitors', *Breast Cancer Research*, 16(4), p. 406. doi: 10.1186/s13058-014-0406-x.

Leonard, G. D. *et al*. (2003) 'The role of ABC transporters in clinical practice', *The Oncologist*, 8(5), pp. 411–24.

Lerner, M. R. *et al.* (1981) 'Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus', *PNAS*, 78(2), pp. 805–9.

Leśniewska, E. and Boguta, M. (2017) 'Novel layers of RNA polymerase III control affecting tRNA gene transcription in eukaryotes', *Open Biology*, 7(2), p. 170001. doi: 10.1098/rsob.170001.

Li, Y. *et al.* (2000) 'Repression of ribosome and tRNA synthesis in secretion-defective cells is signaled by a novel branch of the cell integrity pathway', *Molecular and Cellular Biology*, 20(11), pp. 3843–51. doi: 10.1128/MCB.20.11.3843-3851.2000.

Li, Y. and Tergaonkar, V. (2014) 'Noncanonical functions of telomerase: implications in telomerase-targeted cancer therapies', *Cancer Research*, 74(6), pp. 1639–44. doi: 10.1158/0008-5472.CAN-13-3568.

Liang, Y. *et al.* (2015) 'Epigenetic Activation of TWIST1 by MTDH Promotes Cancer Stem-like Cell Traits in Breast Cancer', *Cancer Research*, 75(17), pp. 3672–3680. doi: 10.1158/0008-5472.CAN-15-0930.

Lin-Marq, N. and Clarkson, S. G. (1998) 'Efficient synthesis, termination and release of RNA polymerase III transcripts in Xenopusextracts depleted of La protein', *The EMBO Journal*, 17(7), pp. 2033–2041. doi: 10.1093/emboj/17.7.2033.

Lin, C. Y. *et al.* (2012) 'Transcriptional Amplification in Tumor Cells with Elevated c-Myc', *Cell*, 151(1), pp. 56–67. doi: 10.1016/j.cell.2012.08.026.

Liu, S. *et al.* (2014) 'Stem Cell Reports Ar ticle Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterpar ts', *Stem Cell Reports*, 2, pp. 78–91. doi: 10.1016/j.stemcr.2013.11.009.

Livraghi, L. and Garber, J. E. (2015) 'PARP inhibitors in the management of breast cancer: current data and future prospects', *BMC Medicine*, 13(188), pp. 1–16. doi: 10.1186/s12916-015-0425-1.

Lobo, S. M. *et al.* (1992) 'A TBP complex essential for transcription from TATA-less but not TATA-containing RNA polymerase III promoters is part of the TFIIIB fraction', *Cell*, 71(6), pp. 1029–40.

Lobo, S. M. and Hernandez, N. (1989) 'A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter', *Cell*, 58(1), pp. 55–67. doi: 10.1016/0092-8674(89)90402-9.

Low, K. C. and Tergaonkar, V. (2013) 'Telomerase: central regulator of all of the hallmarks of cancer', *Trends in Biochemical Sciences*, 38(9), pp. 426–434. doi: 10.1016/j.tibs.2013.07.001.

Lund, R. J. *et al.* (2017) 'RNA Polymerase III Subunit POLR3G Regulates Specific Subsets of PolyA + and Small RNA Transcriptomes and Splicing in Human Pluripotent Stem Cells', *Stem Cell Reports*, 8(5), pp. 1442–1454. doi: 10.1016/j.stemcr.2017.04.016.

Lundgren, K. *et al*. (2009) 'Hypoxia, Snail and incomplete epithelial–mesenchymal transition in breast cancer', *British Journal of Cancer*, 101, pp. 1769–1781. doi: 10.1038/sj.bjc.6605369.

Lunyak, V. V and Atallah, M. (2011) 'Genomic relationship between SINE retrotransposons, Pol III-Pol II transcription, and chromatin organization: the journey from junk to jewel', *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 89(5), pp. 495–504. doi: 10.1139/o11-046.

Makałowski, W. (2000) 'Genomic scrap yard: how genomes utilize all that junk', *Gene*, 259(1–2), pp. 61–7.

Male, G. *et al.* (2015) 'Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly', *Nature Communications*, 6, p. 7387. doi: 10.1038/ncomms8387.

Manaud, N. *et al.* (1998) 'A chimeric subunit of yeast transcription factor IIIC forms a subcomplex with tau95', *Molecular and Cellular Biology*, 18(6), pp. 3191–200. doi: 10.1128 /MCB.18.6.3191.

Maraia, R. J. *et al*. (1994) 'Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III', *Molecular and Cellular Biology*, 14(3), pp. 2147–58.

Maraia, R. J. and Lamichhane, T. N. (2011) '3′ processing of eukaryotic precursor tRNAs', *Wiley Interdisciplinary Reviews: RNA*, 2(3), pp. 362–375. doi: 10.1002/wrna.64.

Marck, C. *et al.* (1993) 'The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs', *PNAS*, 90(9), pp. 4027–31.

Margottin, F. *et al.* (1991) 'Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C', *Science*, 251(4992), pp. 424–6.

Mariner, P. D. *et al.* (2008) 'Human Alu RNA Is a Modular Transacting Repressor of mRNA Transcription during Heat Shock', *Molecular Cell*, 29(4), pp. 499–509. doi: 10.1016/j.molcel.2007.12.013.

Martignetti, J. A. and Brosius, J. (1993) 'BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element', *PNAS*, 90(24), pp. 11563–7.

Marzouki, N. *et al.* (1986) 'Selective proteolysis defines two DNA binding domains in yeast transcription factor τ', *Nature*, 323(6084), pp. 176–178. doi: 10.1038/323176a0.

Masters, B. S. *et al*. (1987) 'Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7', *Cell*, 51(1), pp. 89–99.

Matera, A. G. and Ward, D. C. (1992) 'Oligonucleotide probes for the analysis of specific repetitive DNA sequences by fluorescence in situ hybridization', *Human Molecular Genetics*, 1(7), pp. 535–9.

Matsumoto, K. and Bay, B.-H. (2005) 'Significance of the Y-box proteins in human cancers', *Journal of molecular and genetic medicine : an international journal of biomedical research*, 1(1), pp. 11–7.

Matsuzaki, H. *et al*. (1994) 'Analysis of RNA Chain Elongation and Termination by Saccharomyces cerevisiae RNA Polymerase III', *Journal of Molecular Biology*, 235(4), pp. 1173–1192. doi: 10.1006/jmbi.1994.1072.

Mattaj, I. W. *et al.* (1988) 'Changing the RNA polymerase specificity of U snRNA gene promoters', *Cell*, 55(3), pp. 435–442. doi: 10.1016/0092-8674(88)90029-3.

Mazabraud, A. *et al.* (1987) 'Structure and transcription termination of a lysine tRNA gene from Xenopus laevis', *Journal of Molecular Biology*, 195(4), pp. 835–45.

Mcdermott, S. P. and Wicha, M. S. (2010) 'Targeting breast cancer stem cells'. doi: 10.1016/j.molonc.2010.06.005.

McStay, B. and Grummt, I. (2008) 'The Epigenetics of rRNA Genes: From Molecular to Chromosome Biology', *Annual Review of Cell and Developmental Biology*, 24(1), pp. 131–157. doi: 10.1146/annurev.cellbio.24.110707.175259.

Mergny, J.-L. *et al.* (2005) 'Thermal difference spectra: a specific signature for nucleic acid structures', *Nucleic Acids Research*, 33(16), pp. e138–e138. doi: 10.1093/nar/gni134.

Mertens, C. and Roeder, R. G. (2008) 'Different Functional Modes of p300 in Activation of RNA Polymerase III Transcription from Chromatin Templates', *MOLECULAR AND CELLULAR BIOLOGY*, 28(18), pp. 5764–5776. doi: 10.1128/MCB.01262-07.

Meyer, M. J. *et al.* (2010) 'CD44posCD49fhiCD133/2hi defines xenograft-initiating cells in estrogen receptor-negative breast cancer', *Cancer Research*, 70(11), pp. 4624–33. doi: 10.1158 /0008-5472.CAN-09-3619.

Miki, Y. *et al.* (1994) 'A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1', *Science*, 266(5182), pp. 66–71.

Miller, J. *et al*. (1985) 'Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes', *The EMBO Journal*, 4(6), pp. 1609–14.

Mizuta, K. and Warner, J. R. (1994) 'Continued functioning of the secretory pathway is essential for ribosome synthesis', *Molecular and Cellular Biology*, 14(4), pp. 2493–502.

Moinfar, F. *et al.* (2003) 'Androgen receptors frequently are expressed in breast carcinomas: potential relevance to new therapeutic strategies', *Cancer*, 98(4), pp. 703–11. doi: 10.1002/cncr.11532.

Moitra, K. and Karobi (2015) 'Overcoming Multidrug Resistance in Cancer Stem Cells', *BioMed Research International*. Hindawi, 2015, pp. 1–8. doi: 10.1155/2015/635745.

Montijn, M. B. *et al.* (1999) 'The 5S rRNA gene clusters have a defined orientation toward the nucleolus in Petunia hybrida and Crepis capillaris', *Chromosome Research*, 7(5), pp. 387–99.

Moqtaderi, Z. *et al.* (2010) 'Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells', *Nature Structural & Molecular Biology*, 17(5), pp. 635–40. doi: 10.1038/nsmb.1794.

Moqtaderi, Z. and Struhl, K. (2004) 'Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes', *Molecular and Cellular Biology*, 24(10), pp. 4118–27.

Morel, A.-P. *et al.* (2008) 'Generation of breast cancer stem cells through epithelial-mesenchymal transition', *PloS ONE*. Public Library of Science, 3(8), p. e2888. doi: 10.1371/journal.pone.0002888.

Morikawa, A. and Henry, N. L. (2015) 'Palbociclib for the Treatment of Estrogen Receptor-Positive, HER2-Negative Metastatic Breast Cancer', *Clinical Cancer Research*, 21(16), pp. 3591–3596. doi: 10.1158/1078-0432.CCR-15-0390.

Moss, W. N. *et al.* (2014) 'RNA families in Epstein-Barr virus', *RNA biology*, 11(1), pp. 10–7. doi: 10.4161/rna.27488.

Movahedi, A. *et al.* (2015) 'RNA-directed DNA methylation in plants', *Plant Cell Reports*, 34(11), pp. 1857–1862. doi: 10.1007/s00299-015-1839-0.

Murphy, S. *et al.* (1992) 'Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes', *Molecular and Cellular Biology*, 12(7), pp. 3247–61.

Murphy, S. *et al*. (1986) 'A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes', *Nucleic Acids Research*, 14(23), pp. 9243–60.

Mus, E. *et al*. (2007) 'Dendritic BC200 RNA in aging and in Alzheimer's disease', *PNAS*, 104(25), pp. 10679–10684. doi: 10.1073/pnas.0701532104.

Naldini, L. *et al.* (1996) 'Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector', *PNAS*, 93(21), pp. 11382–8.

Neumeister, V. *et al.* (2010) 'In Situ Identification of Putative Cancer Stem Cells by Multiplexing ALDH1, CD44, and Cytokeratin Identifies Breast Cancer Patients with Poor Prognosis', *The American Journal of Pathology*, 176(5), pp. 2131–2138. doi: 10.2353/ajpath.2010.090712.

Neve, R. M. *et al.* (2006) 'A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes', *Cancer cell*, 10(6), pp. 515–27. doi: 10.1016/j.ccr.2006.10.008.

Nguyen, V. T. *et al.* (2001) '7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes', *Nature*, 414(6861), pp. 322–325. doi: 10.1038/35104581.

Ni, M. *et al.* (2011) 'Targeting Androgen Receptor in Estrogen Receptor-Negative Breast Cancer', *Cancer Cell*, 20(1), pp. 119–131. doi: 10.1016/j.ccr.2011.05.026.

Nielsen, S., *et al*. (2013) 'Mechanism of Eukaryotic RNA Polymerase III Transcription Termination', *Science*, 340(6140), pp. 1577–1580. doi: 10.1126/science.1237934.

Nikolov, D. B. *et al.* (1996) 'Crystal structure of a human TATA box-binding protein/TATA element complex', *Cell Biology*, 93, pp. 4862–4867.

Noma, K. *et al.* (2006) 'A Role for TFIIIC Transcription Factor Complex in Genome Organization', *Cell*, 125(5), pp. 859–872. doi: 10.1016/j.cell.2006.04.028.

O'Brien, C. A. and Wolin, S. L. (1994) 'A possible role for the 60-kD Ro autoantigen in a discard pathway for defective 5S rRNA precursors', *Genes & Development*, 8(23), pp. 2891–903.

Oettel, S. *et al.* (1997) 'Human transcription factors IIIC2 , IIIC1 and a novel component IIIC0 fulfil different aspects of DNA binding to various pol III genes', *Nucleic Acids Research*, 25(12), pp. 2440–7.

Oettel, S., Kober, I. and Seifart, K. H. (1998) 'The activity binding to the termination region of several pol III genes represents a separate entity and is distinct from a novel component enhancing U6 snRNA transcription', *Nucleic Acids Research*, 26(19), pp. 4324–31.

Oler, A. J. *et al.* (2010) 'Human RNA Polymerase III transcriptomes and relationships to Pol II promoters , enhancer-binding factors and chromatin domains', *Nature Structural & Molecular Biology*, 17(5), pp. 620–628.

Olmeda, D. *et al.* (2007) 'SNAI1 Is Required for Tumor Growth and Lymph Node Metastasis of Human Breast Carcinoma MDA-MB-231 Cells', *Cancer Res*, 67(24), pp. 11721–31. doi: 10.1158/0008-5472.CAN-07-2318.

Orphanides, G. *et al*. (1996) 'The general transcription factors of RNA polymerase II', *Genes & Development*, 10(21), pp. 2657–83.

Oualla, K. *et al.* (2017) 'Novel therapeutic strategies in the treatment of triple-negative breast cancer', *Therapeutic Advances in Medical Oncology*, 9(7), pp. 493–511. doi: 10.1177/1758834017711380.

Owens, T. W. and Naylor, M. J. (2013) 'Breast cancer stem cells', *Frontiers in Physiology*, 4 AUG(August), pp. 1–10. doi: 10.3389/fphys.2013.00225.

Pattabiraman, D. R. and Weinberg, R. A. (2014) 'Tackling the cancer stem cells — what challenges do they pose?', *Nature Reviews Drug Discovery*, 13. doi: 10.1038/nrd4253.

Paula, A. da C. and Lopes, C. (2017) 'Implications of Different Cancer Stem Cell Phenotypes in Breast Cancer', *Anticancer Research*, 37(5), pp. 2173–2183. doi: 10.21873/anticanres.11552.

Pavon-Eternod, M. *et al.* (2009) 'tRNA over-expression in breast cancer and functional consequences', *Nucleic Acids Research*, 37(21), pp. 7268–80. doi: 10.1093/nar/gkp787.

Pavon-Eternod, M. *et al.* (2013) 'Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells', *RNA*, 19(4), pp. 461–6. doi: 10.1261/rna.037507.112.

Peitzsch, C. *et al.* (2013) 'Discovery of the cancer stem cell related determinants of radioresistance', *Radiotherapy and Oncology, 108, pp. 378–387. doi: 10.1016/j.radonc.2013.06.003.*

Perou, C. M. et al. (2000) 'Molecular portraits of human breast tumours', *Nature*, 406(6797), pp. 747–752. doi: 10.1038/35021093.

Phillips, T. M. *et al*. (2006) 'The Response of CD24-/low/CD44+ Breast Cancer-Initiating Cells to Radiation', *Journal of the National Cancer Institute*, 98(24), pp. 1777–1785. doi: 10.1093/jnci/djj495.

Pinatel, E. M. et al. (2014) 'miR-223 is a coordinator of breast cancer progression as revealed by bioinformatics predictions', *PloS ONE*, 9(1), p. e84859. doi: 10.1371/journal.pone.0084859.

Ponicsan, S. L. *et al*. (2010) 'Genomic gems: SINE RNAs regulate mRNA production', *Current opinion in genetics & development*, 20(2), pp. 149–55. doi: 10.1016/j.gde.2010.01.004.

Powers, T. and Walter, P. (1999) 'Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in Saccharomyces cerevisiae', *Molecular biology of the cell*, 10(4), pp. 987–1000.

Prat, A. and Perou, C. M. (2011) 'Deconstructing the molecular portraits of breast cancer', *Molecular Oncology*, 5(1), pp. 5–23. doi: 10.1016/j.molonc.2010.11.003.

Pylayeva-Gupta, Y. *et al*. (2011) 'RAS oncogenes: weaving a tumorigenic web', *Nature Reviews Cancer*, 11(11), pp. 761–774. doi: 10.1038/nrc3106.

Raha, D. et al. (2010) 'Close association of RNA polymerase II and many transcription factors with Pol III genes', *PNAS*, 107(8), pp. 3639–44. doi: 10.1073/pnas.0911315106.

Ream, T. S. et al. (2009) 'Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II', *Molecular Cell*, 33(2), pp. 192–203. doi: 10.1016/j.molcel.2008.12.015.

Reich, P. R. *et al*. (1966) 'RNA of low molecular weight in KB cells infected with adenovirus type 2', *Journal of Molecular Biology*, 17(2), pp. 428–39.

Reina, J. H. et al. (2006) 'Maf1, a New Player in the Regulation of Human RNA Polymerase III Transcription', *PLoS ONE*, 1(1), p. e134. doi: 10.1371/journal.pone.0000134.

Reiner, R. et al. (2006) 'A role for the catalytic ribonucleoprotein RNase P in RNA polymerase III transcription', *Genes & Development*, 20,pp. 1621–1635. doi: 10.1101/gad.386706.

Reiner, R. et al. (2008) 'Function and Assembly of a Chromatin-Associated RNase P that Is Required for Efficient Transcription by RNA Polymerase I', *PloS ONE*, 3(12), p. e4072.

Reis-Filho, J. S. and Tutt, A. N. J. (2007) 'Triple negative tumours: a critical review', *Histopathology*, 52(1), pp. 108–118. doi: 10.1111/j.1365-2559.2007.02889.

Renaud, M. et al. (2014) 'Gene duplication and neofunctionalization: POLR3G and POLR3GL', *Genome Research*, 24(1), pp. 37–51. doi: 10.1101/gr.161570.113.

Reya, T. et al. (2001) 'Stem cells, cancer, and cancer stem cells', *Nature*, 414(6859), pp. 105–111. doi: 10.1038/35102167.

Richards, K. L. et al. (2009) 'Genome-Wide Hypomethylation in Head and Neck Cancer Is More Pronounced in HPV-Negative Tumors and Is Associated with Genomic Instability', *PLoS ONE*, 4(3), p. e4941. doi: 10.1371/journal.pone.0004941.

Rijal, K. and Maraia, R. J. (2016) 'Active Center Control of Termination by RNA Polymerase III and tRNA Gene Transcription Levels In Vivo', *PLoS genetics*, 12(8), p. e1006253. doi: 10.1371/journal.pgen.1006253.

Roeder, R. G. and Rutter, W. J. (1969) 'Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms', *Nature*, 224(5216), pp. 234–237.

Roy, K. et al. (2016) 'Common genomic elements promote transcriptional and DNA replication roadblocks', *Genome Research*, 26(10), pp. 1363–1375. doi: 10.1101/gr.204776.116.

Ruet, A. et al. (1984) 'Isolation of a class C transcription factor which forms a stable complex with tRNA genes', *The EMBO Journal*, 3(2), pp. 343–50.

Rüth, J. et al. (1996) 'A suppressor of mutations in the class III transcription system encodes a component of yeast TFIIIB', *The EMBO Journal*, 15(8), pp. 1941–9.

Sadowski, C. L. et al. (1996) 'The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein', *PNAS*, 93(9), pp. 4289–93.

Sakonju, S. *et al*. (1980) 'A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region', *Cell*, 19(1), pp. 13–25. doi: 10.1016/0092-8674(80)90384-0.

Salvati, E. et al. (2014) 'Evidence for G-quadruplex in the promoter of vegfr-2 and its targeting to inhibit tumor angiogenesis', *Nucleic Acids Research*, 42(5), pp. 2945–2957. doi: 10.1093/nar/gkt1289.

Schaub, M. et al. (1997) 'Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III', *The EMBO Journal*, 16(1), pp. 173–81. doi: 10.1093/emboj/16.1.173.

Schmitt, M. E. and Clayton, D. A. (1993) 'Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in Saccharomyces cerevisiae', *Molecular and Cellular Biology*, 13(12), pp. 7935–41.

Schramm, L. et al. (2000) 'Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters', *Genes & Development*, 14(20), pp. 2650–63.

Schramm, L. and Hernandez, N. (2002) 'Recruitment of RNA polymerase III to its target promoters', *Genes & Development*, 16(20), pp. 2593–620. doi: 10.1101/gad.1018902.

Schubbert, S. *et al*. (2007) 'Hyperactive Ras in developmental disorders and cancer', *Nature Reviews Cancer*, 7(4), pp. 295–308. doi: 10.1038/nrc2109.

Schulman, D. B. and Setzer, D. R. (2002) 'Identification and characterization of transcription factor IIIA from Schizosaccharomyces pombe', *Nucleic Acids Research*, 30(13), pp. 2772–81.

Schultz, P. et al. (1989) 'The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy', *The EMBO Journal*, 8(12), pp. 3815–24.

Schwartz, L. B. et al. (1974) 'Isolation and partial characterization of the multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in the mouse myeloma, MOPC 315', *The Journal of Biological Chemistry*, 249(18), pp. 5889–97.

Serruya, R. et al. (2015) 'Human RNase P ribonucleoprotein is required for formation of initiation complexes of RNA polymerase III', *Nucleic Acids Research*, 43(11), pp. 5442–5450. doi: 10.1093/nar/gkv447.

Sharp, S. et al. (1981) 'Internal control regions for transcription of eukaryotic tRNA genes', *PNAS*, 78(11), pp. 6657–61.

Sheridan, C. et al. (2006) 'CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis', *Breast Cancer Research*, 8(5), p. R59. doi: 10.1186/bcr1610.

Shivaswamy, S. et al. (2008) 'Dynamic Remodeling of Individual Nucleosomes Across a Eukaryotic Genome in Response to Transcriptional Perturbation', *PLoS Biology*, 6(3), p. e65. doi: 10.1371/journal.pbio.0060065.

Shultz, L. D. et al. (2005) 'Human Lymphoid and Myeloid Cell Development in NOD/LtSz-scid IL2R null Mice Engrafted with Mobilized Human Hemopoietic Stem Cells', *The Journal of Immunology*, 174(10), pp. 6477–6489. doi: 10.4049/jimmunol.174.10.6477.

Siddiqui-Jain, A. et al. (2002) 'Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription', *PNAS*, 99(18), pp. 11593–11598. doi: 10.1073/pnas.182256799.

Siegel, R. *et al*. (2013) 'Cancer statistics, 2013', *CA: A Cancer Journal for Clinicians*, 63(1), pp. 11–30. doi: 10.3322/caac.21166.

Sim, S. and Wolin, S. L. (2011) 'Emerging roles for the Ro 60-kDa autoantigen in noncoding RNA metabolism', *Wiley Interdisciplinary Reviews: RNA*, 2(5), pp. 686–699. doi: 10.1002/wrna.85.

Simonsson, T. *et al*. (1998) 'DNA tetraplex formation in the control region of c-myc', *Nucleic Acids Research*, 26(5), pp. 1167-72

Singh, A. and Settleman, J. (2010) 'EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer', *Oncogene*, 29, pp. 4741-51. doi: 10.1038/onc.2010.215.

Singh, R. et al. (2016) 'Regulation of alternative splicing of Bcl-x by BC200 contributes to breast cancer pathogenesis', *Cell Death & Disease*, 7(6), p. e2262. doi: 10.1038/cddis.2016.168.

Singh, S. et al. (2013) 'Aldehyde dehydrogenases in cellular responses to oxidative/ electrophilicstress', *Free Radical Biology and Medicine*, 56, pp. 89–101. doi: 10.1016 /j.freeradbiomed.2012.11.010.

Sinn, E. et al. (1995) 'Cloning and characterization of a TFIIIC2 subunit (TFIIIC beta) whose presence correlates with activation of RNA polymerase III-mediated transcription by adenovirus E1A expression and serum factors', *Genes & Development*, 9(6), pp. 675–85.

Soares, A. R. and Santos, M. (2017) 'Discovery and function of transfer RNA-derived fragments and their role in disease', *Wiley Interdisciplinary Reviews: RNA*, pp. 1–13. doi: 10.1002 /wrna.1423.

Söderlund, H. et al. (1976) 'A new species of virus-coded low molecular weight RNA from cells infected with adenovirus type 2', *Cell*, 7(4), pp. 585–93.

Soerensen, P. D. and Frederiksen, S. (1991) 'Characterization of human 5S rRNA genes', *Nucleic Acids Research*, 19(15), pp. 4147–51.

Soragni, E. and Kassavetis, G. A. (2008) 'Absolute Gene Occupancies by RNA Polymerase III, TFIIIB, and TFIIIC in Saccharomyces cerevisiae', *Journal of Biological Chemistry*, 283(39), pp. 26568–76. doi: 10.1074/jbc.M803769200.

Sorlie, T. et al. (2003) 'Repeated observation of breast tumor subtypes in independent gene expression data sets', *PNAS*, 100(14), pp. 8418–23. doi: 10.1073/pnas.0932692100.

Sørlie, T. et al. (2001) 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *PNAS*, 98(19), pp. 10869–74. doi: 10.1073/pnas.191367098.

Soule, H. D. et al. (1973) 'A Human Cell Line From a Pleural Effusion Derived From a Breast Carcinoma', *Journal of the National Cancer Institute*, 51(5), pp. 1409–1416. doi: 10.1093/jnci/51.5.1409.

Soule, H. D. et al. (1990) 'Isolation and Characterization of a Spontaneously Immortalized Isolation and Characterization of a Spontaneously Immortalized Human Breast', *Cancer*, 50(18), pp. 6075–6086.

Stadler, P. F. et al. (2009) 'Evolution of Vault RNAs', *Molecular Biology and Evolution*, 26(9), pp. 1975–1991. doi: 10.1093/molbev/msp112.

Stefano, J. E. (1984) 'Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts', *Cell*, 36(1), pp. 145–54.

Steffensen, D. M. *et al*. (1974) 'Localisation of 5S ribosomal RNA genes on human chromosome 1', *Nature*, 252(5485), pp. 741–3.

Steiner, E. et al. (2006) 'Cellular functions of vaults and their involvement in multidrug resistance', *Current Drug Targets*, 7(8), pp. 923–34.

Sun, D. et al. (2005) 'Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents', *Nucleic Acids Research*, 33(18), pp. 6070–80. doi: 10.1093/nar/gki917.

Sundquist, W. I. and Klug, A. (1989) 'Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops', *Nature*, 342(6251), pp. 825–829. doi: 10.1038/342825a0.

Suprenant, K. A. (2002) 'Vault ribonucleoprotein particles: sarcophagi, gondolas, or safety deposit boxes?', *Biochemistry*, 41(49), pp. 14447–54.

Sutcliffe, J. E. et al. (1999) 'RNA polymerase III transcription factor IIIB is a target for repression by pocket proteins p107 and p130', *Molecular and Cellular Biology*, 19(6), pp. 4255–61. doi: 10.1128/MCB.19.6.4255.

Swanson, R. N. et al. (1991) 'Isolation of TFC1, a gene encoding one of two DNA-binding subunits of yeast transcription factor tau (TFIIIC)', *PNAS*, 88(11), pp. 4887–91.

Takada, S. et al. (2000) 'A TRF1:BRF complex directs Drosophila RNA polymerase III transcription', *Cell*, 101(5), pp. 459–69.

Tanaka, M., *et al*. (1992) 'Promoter-selective activation domains in Oct-1 and Oct-2 direct differential activation of an snRNA and mRNA promoter', *Cell*, 68(4), pp. 755–67.

Tang, R.-B. et al. (2005) 'Increased level of polymerase III transcribed Alu RNA in hepatocellular carcinoma tissue', *Molecular carcinogenesis*, 42(2), pp. 93–6. doi: 10.1002/mc.20057.

Teichmann, M. et al. (1997) 'Functional interchangeability of TFIIIB components from yeast and human cells in vitro', *The EMBO Journal*, 16(15), pp. 4708–16. doi: 10.1093/emboj /16.15.4708.

Teichmann, M. and Seifart, K. H. (1995) 'Physical separation of two different forms of human TFIIIB active in the transcription of the U6 or the VAI gene in vitro', *The EMBO Journal*, 14(23), pp. 5974–83.

Tiede, B. and Kang, Y. (2011) 'From milk to malignancy: the role of mammary stem cells in development, pregnancy and breast cancer', *Cell Research*, 21(2), pp. 245-57. doi: 10.1038/cr.2011.11.

Thompson, E. W. and Haviv, I. (2011) 'The social aspects of EMT-MET plasticity', *Nature Medicine*, 17(499), pp. 1048–49. doi: 10.1038/nm.2437.

Thuillier, V. et al. (1995) 'A mutation in the C31 subunit of Saccharomyces cerevisiae RNA polymerase III affects transcription initiation', *The EMBO Journal*, 14(2), pp. 351–9.

Tiranti, V. et al. (1997) 'Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database', *Human Molecular Genetics*, 6(4), pp. 615–25.

Ullu, E. and Tschudi, C. (1984) 'Alu sequences are processed 7SL RNA genes', *Nature*, 312(5990), pp. 171–2.

Ullu, E. and Weiner, A. M. (1985) 'Upstream sequences modulate the internal promoter of the human 7SL RNA gene', *Nature*, 318(6044), pp. 371–4.

Upadhya, R. *et al*. (2002) 'Maf1 is an essential mediator of diverse signals that repress RNA polymerase III transcription', *Molecular Cell*, 10(6), pp. 1489–94.

Vaillant, F. et al. (2008) 'The Mammary Progenitor Marker CD61/ 3 Integrin Identifies Cancer Stem Cells in Mouse Models of Mammary Tumorigenesis', *Cancer Research*, 68(19), pp. 7711–7717. doi: 10.1158/0008-5472.CAN-08-1949.

Valentijn, L. J. et al. (2012) 'Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification', *PNAS*, 109(47), pp. 19190–5. doi: 10.1073/pnas.1208215109.

Valenzuela, L. *et al*. (2009) 'Transcription Independent Insulation at TFIIIC-Dependent Insulators', *Genetics*, 183(1), pp. 131–148. doi: 10.1534/genetics.109.106203.

Vannini, A. and Cramer, P. (2012) 'Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries', *Molecular Cell*, 45(4), pp. 439–446. doi: 10.1016/j.molcel.2012.01.023.

Waldschmidt, R. *et al*. (1991) 'Identification of transcription factors required for the expression of mammalian U6 genes in vitro', *The EMBO Journal*, 10(9), pp. 2595–603.

Walz, S. et al. (2014) 'Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles', *Nature*, 511(7510), pp. 483–7. doi: 10.1038/nature13473.

Wang, H. et al. (2002) 'Dendritic BC1 RNA: functional role in regulation of translation initiation', *Journal of Neuroscience*, 22(23), pp. 10232–41.

Wang, X. et al. (2016) 'PAF-Wnt signaling-induced cell plasticity is required for maintenance of breast cancer cell stemness', *Nature Communications*, 7, p. 10633. doi: 10.1038/ncomms10633.

Wang, Y. et al. (2013) 'The Role of Snail in EMT and Tumorigenesis', *Current Cancer Drug Targets*, 13(9), pp. 963–972.

Wang, Z. et al. (2000) 'Nuclear factor 1 (NF1) affects accurate termination and multiple-round transcription by human RNA polymerase III', *The EMBO Journal*, 19(24), pp. 6823–32. doi: 10.1093/emboj/19.24.6823.

Wang, Z. et al. (2008) 'Combinatorial patterns of histone acetylations and methylations in the human genome', *Nature Genetics*, 40(7), pp. 897–903. doi: 10.1038/ng.154.

Wang, Z. and Roeder, R. G. (1997) 'Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation', *Genes & Development*, 11(10), pp. 1315–26.

Wang, Z. and Roeder, R. G. (1998) 'DNA topoisomerase I and PC4 can interact with human TFIIIC to promote both accurate termination and transcription reinitiation by RNA polymerase III', *Molecular Cell*, 1(5), pp. 749–57.

Watson, J. B. and Sutcliffe, J. G. (1987) 'Primate Brain-Specific Cytoplasmic Transcript of the Alu Repeat Family', *Molecular and Cellular Biology*, 7(9), pp. 3324–3327.

Werner, F. (2008) 'Structural evolution of multisubunit RNA polymerases', Trends in microbiology, 16(6), pp. 247–250. doi: https://doi.org/10.1016/j.tim.2008.03.008.

Werner, M. et al. (1993) 'Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB', The Journal of Biological Chemistry, 268(28), pp. 20721–4.

Weser, S. et al. (2004) 'Transcription Factor (TF)-like Nuclear Regulator, the 250-kDa Form of Homo sapiens TFIIIB″, Is an Essential Component of Human TFIIIC1 Activity', Journal of Biological Chemistry, 279(26), pp. 27022–27029. doi: 10.1074/jbc.M312790200.

White, R. J. et al. (1995) 'Cell cycle regulation of RNA polymerase III transcription', *Molecular and Cellular Biology*, 15(12), pp. 6653–62.

White, R. J. et al. (1996) 'Repression of RNA polymerase III transcription by the retinoblastoma protein', *Nature*, 382(6586), pp. 88–90. doi: 10.1038/382088a0.

White, R. J. (2004) 'RNA polymerase III transcription and cancer', *Oncogene*, 23(18), pp. 3208–16. doi: 10.1038/sj.onc.1207547.

White, R. J. (2011) 'Transcription by RNA polymerase III: more complex than we thought', *Nature reviews Genetics*, 12(7), pp. 459–63. doi: 10.1038/nrg3001.

White, R. J. and Jackson, S. P. (1992) 'Mechanism of TATA-binding protein recruitment to a TATA-less class III promoter', *Cell*, 71(6), pp. 1041–53.

Wicker, T. et al. (2007) 'A unified classification system for eukaryotic transposable elements', *Nature Reviews Genetics*, 8(12), pp. 973–982. doi: 10.1038/nrg2165.

Wierzbicki, A. T. *et al*. (2008) 'Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes', *Cell*, 135(4), pp. 635–648. doi: 10.1016/j.cell.2008.09.035.

Wild, T. and Cramer, P. (2012) 'Biogenesis of multisubunit RNA polymerases', *Trends in Biochemical Sciences*, 37(3), pp. 99–105. doi: 10.1016/j.tibs.2011.12.001.

Winter, A. G. et al. (2000) 'RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors', *PNAS*, 7(23), pp. 12619–24.

Woiwode, A. et al. (2008) 'PTEN represses RNA polymerase III-dependent transcription by targeting the TFIIIB complex', *Molecular and Cellular Biology*, 28(12), pp. 4204–14. doi: 10.1128/MCB.01912-07.

Wong, M. W. et al. (1998) 'The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1', *Molecular and Cellular Biology*, 18(1), pp. 368–77.

Wong, R. C.-B. et al. (2011) 'A novel role for an RNA polymerase III subunit POLR3G in regulating pluripotency in human embryonic stem cells', *Stem Cells*, 29(10), pp. 1517–27. doi: 10.1002/stem.714.

Wooster, R. et al. (1995) 'Identification of the breast cancer susceptibility gene BRCA2', *Nature*, 378(6559), pp. 789–792. doi: 10.1038/378789a0.

Wright, M. H. et al. (2008) 'Brca1 breast tumors contain distinct CD44+/CD24- and CD133+cells with cancer stem cell characteristics', *Breast Cancer Research*, 10(1), p. R10. doi: 10.1186/bcr1855.

Wu, C. C. et al. (2012) 'RNA polymerase III subunit architecture and implications for open promoter complex formation', *PNAS*, 109(47), pp. 19232–19237. doi: DOI 10.1073/pnas.1211665109.

Wu, L. H. et al. (2013) 'Decoy oligonucleotide rescues IGF1R expression from MicroRNA-223 suppression', *PloS ONE*, 8(12), p. e82167. doi: 10.1371/journal.pone.0082167.

Xiang, S. et al. (2010) 'Methylation status of individual CpG sites within Alu elements in the human genome and Alu hypomethylation in gastric carcinomas', *BMC cancer*, 10, p. 44. doi: 10.1186/1471-2407-10-44.

Xie, H. et al. (2009) 'High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum', *Nucleic Acids Research*, 37(13), pp. 4331–4340. doi: 10.1093/nar/gkp393.

Xu, H. et al. (2017) 'CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours', *Nature Communications*, 8, p. 14432. doi: 10.1038/ncomms14432.

Xu, Y. and Sugiyama, H. (2006) 'Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb)', *Nucleic Acids Research*, 34(3), pp. 949–954. doi: 10.1093/nar/gkj485.

Yanagawa, M. et al. (2012) 'Luminal A and luminal B (HER2 negative) subtypes of breast cancer consist of a mixture of tumors with different genotype Masumi', *BioMed Central Research Notes*, 5, p. 376. doi: 10.1186/1756-0500-5-376.

Yang, F. et al. (2017) 'Breast cancer stem cell: the roles and therapeutic implications', *Cellular and Molecular Life Sciences*, 74, pp. 951–6. doi: 10.1007/s00018-016-2334-7.

Yersal, O. and Barutca, S. (2014) 'Biological subtypes of breast cancer: Prognostic and therapeutic implications', *World Journal of Clinical Oncology*, 5(3), pp. 412–24. doi: 10.5306 /wjco.v5.i3.412.

Yik, J. H. N. et al. (2003) 'Inhibition of P-TEFb (CDK9/cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA', *Molecular Cell*, 12(4), pp. 971–982. doi: 10.1016/S1097-2765(03)00388-5.

Yoon, J. B. et al. (1995) 'Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes', *Molecular and Cellular Biology*, 15(4), pp. 2019–27.

Yoon, J. B. and Roeder, R. G. (1996) 'Cloning of two proximal sequence element-binding transcription factor subunits (gamma and delta) that are required for transcription of small nuclear RNA genes by RNA polymerases II and III and interact with the TATA-binding protein', *Molecular and Cellular Biology*, 16(1), pp. 1–9.

Yoshinaga, S. K. *et al*. (1987) 'Resolution of human transcription factor TFIIIC into two functional components', *PNAS*, 84(11), pp. 3585–9.

Young, L. S. and Rickinson, A. B. (2004) 'Epstein-Barr virus: 40 years on', *Nature Reviews Cancer*, 4, pp. 757–768. doi: 10.1038/nrc1452.

Yuan, G.-C. et al. (2005) 'Genome-Scale Identification of Nucleosome Positions in S. cerevisiae', *Science*, 309(5734), pp. 626–630. doi: 10.1126/science.1112178.

Yustein, J. T. et al. (2010) 'Induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B-cell model', *PNAS*, 107(8), pp. 3534–9. doi: 10.1073/pnas.0901230107.

Zambetti, G. P. et al. (1992) 'Wild-type p53 mediates positive regulation of gene expression through a specific DNA sequence element', *Genes & Development*, 6(7), pp. 1143–52. doi: 10.1101/GAD.6.7.1143.

Zaragoza, D. et al. (1998) 'Rapamycin induces the G0 program of transcriptional repression in yeast by interfering with the TOR signaling pathway', *Molecular and Cellular Biology*, 18(8), pp. 4463–70. doi: 10.1128/MCB.18.8.4463.

Zhang, H. et al. (2000) 'Regulation of cellular growth by the Drosophila target of rapamycin dTOR', *Genes & Development*, 14(21), pp. 2712–24. doi: 10.1101/GAD.835000.

Zhong, S. *et al*. (2004) 'Epidermal growth factor enhances cellular TATA binding protein levels and induces RNA polymerase I- and III-dependent gene activity', *Molecular and Cellular Biology*, 24(12), pp. 5119–29. doi: 10.1128/MCB.24.12.5119-5129.2004.

Zhou, L. et al. (2010) 'The prognostic role of cancer stem cells in breast cancer: a meta-analysis of published literatures', *Breast Cancer Research and Treatment*, 122(3), pp. 795–801. doi: 10.1007/s10549-010-0999-4.

Zhou, V. W. *et al*. (2011) 'Charting histone modifications and the functional organization of mammalian genomes', *Nature Reviews Genetics*, 12(1), pp. 7–18. doi: 10.1038/nrg2905.

Van Zon, A. et al. (2001) 'Multiple human vault RNAs: Expression and association with the vault complex', *Journal of Biological Chemistry*, 276(40), pp. 37715–37721. doi: 10.1074 /jbc.M106055200.