# Chapter 6

# *In vivo* CRISPR screening to identify tumour cell intrinsic regulators of metastatic colonisation

## 6.1 Introduction

### 6.1.1 What is metastasis?

It is estimated that 90% of cancer-associated deaths are caused by metastasis.[281] Metastasis is a process whereby cancer cells spread from the primary tumour to a secondary organ. Multiple stages are involved: invasion of surrounding tissue, intravasation and survival in the circulatory system, extravasation and proliferation in a distant tissue (in a process known as 'colonisation') (Fig. 6.1). Metastasis is controlled by both tumour cell intrinsic factors and extrinsic factors in the tumour microenvironment (reviewed by Liu, *et al.,* 2017).[282] Many studies have been carried out to identify such factors but we still have a lot to learn about this process, particularly the colonisation step.

### 6.1.2 Metastasis models

Due to the biologically complex nature of this process, animal models are critical for studying metastasis. In particular, the role of the immune system in metastasis makes it difficult to accurately study this process *in vitro*. Research has shown that the immune system can act both to inhibit and promote metastasis at various stages of the cascade. Cytotoxic CD8+ T cells, NK cells and M1-like macrophages have inhibitory roles and will recognise and kill tumour cells.[283] However, tumour cells can develop mechanisms to evade the immune response, for example by downregulating the expression of MHC class I molecules which are recognised by CD8+ T cells.[284] T regulatory cells and M2-like macrophages can also help to promote metastasis by inhibiting CD8+ T cells and NK cells.[283] Immune cells play various roles throughout the metastatic cascade (as reviewed by Blomberg, *et al*., 2018).[285] In the early stages of

dissemination from the primary tumour, immune cells regulate the extracellular matrix, formation of blood and lymph vessels, and can promote migration. After intravasation, circulating tumour cells are under further attack by NK cells and T cells and must adapt to evade these responses. However, some immune cells can aid survival and assist with extravasation to distant sites. For example, interaction with platelets in the circulation can shield tumour cells from NK cells[286,287] and can also promote an invasive phenotype in the tumour cells.[288] If cells survive these pressures, they must then establish another immunosuppressive environment at the secondary site to allow further growth. The interaction between tumour cells and the host environment, particularly the immune system is pivotal in determining the success of metastasis. Identifying proteins that are up- or down-regulated in tumour cells to allow them to metastasise is vital for developing therapies to prevent this.

One of the most common *in vivo* strategies is to transplant cancer cell lines or tissues into mice.[289] This can be done using cells/tissues derived from an animal with the same genetic background (syngeneic) or using human cells/tissues in immunocompromised mice (xenograft). The interaction between the tumour and the host is critical in metastasis. The syngeneic model is ideal for studying this as the tumour cells are from the same species as the host, whereas human xenograft models may face species-specificity issues.

Two experimental approaches are used with transplantable models: experimental metastasis assays (injection of tumour cells directly into the circulation) and spontaneous metastasis assays (injection of tumour cells into a tissue from which they then have to enter the circulation). Experimental metastasis assays allow analysis of the later stages of metastasis (extravasation and colonisation), whereas spontaneous assays follow all stages from the primary site. The spontaneous model is more clinically relevant but the experimental model is quicker and avoids issues with the primary tumour mass growing too large (and the mouse having to be sacrificed) before the cells have had time to enter the circulation and metastasise.

One of the most commonly used experimental models involves intravenous injection of B16 mouse melanoma cells into the tail vein, typically leading to pulmonary metastases.[290] B16 tumour cells were derived from a spontaneous melanoma that occurred in the C57BL/6 mouse strain. Several derivatives of this line have been generated with varying metastatic potential and sites of colonisation.[291,292] This model (using B16-F10 cells; a highly metastatic derivative[290]) has been used extensively in the Adams' lab.[293]

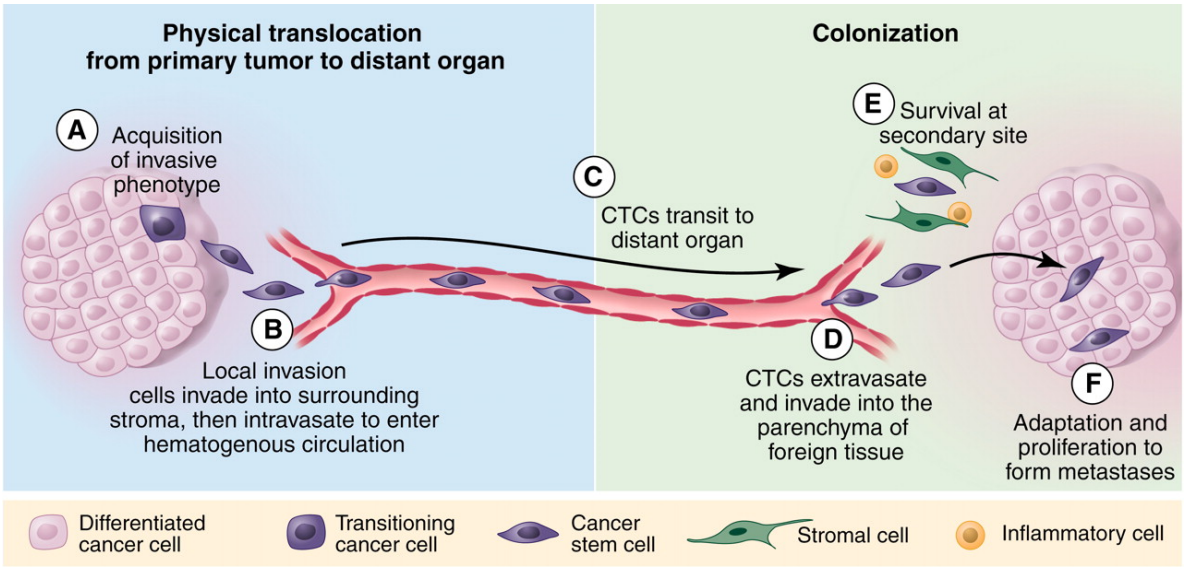


**Figure 6.1. Metastatic cascade.** Tumour cells acquire an invasive phenotype **(a)**, begin to invade surrounding tissue and then intravasate into the vasculature **(b)**. Circulating tumour cells (CTCs) travel through the circulatory system **(c)** until they extravasate into a distant tissue **(d)**. Here, they must survive challenges such as the innate immune system **(e)** and adapt to the new microenvironment, allowing them proliferate and form metastases **(f)**. Figure taken from [281].

### 6.1.3 Screening for regulators of metastasis

By applying high-throughput screening strategies to established *in vivo* models, we can interrogate the metastatic process in a systematic and unbiased manner. A recent study identified tumour extrinsic ('host') factors that regulate metastasis by carrying out an experimental metastasis assay in 810 genetically-modified mouse lines.[293] Tumour intrinsic factors have also been identified by applying functional genomic approaches, such as shRNA and cDNA libraries, to cells *in vitro* then observing the phenotypic effect in *in vivo* metastasis models.[294-296] More recently, CRISPR/Cas9 technology has been used for various loss-of-function screens *in vivo,*[297-299] including the study of regulators of tumour growth and metastasis in mice.[300]

### 6.1.4 CRISPR activation

Originally a tool developed to allow for genome editing[87,88], CRISPR/Cas9 has been repurposed for many applications, including gene regulation. A catalytically inactive version of Cas9 (dCas9) was engineered by introducing point mutations into the RuvC1 and HNH nuclease domains.[96] dCas9 is unable to cleave DNA but can still be targeted to a genomic region by a gRNA. By fusing transcriptional regulation domains to dCas9, the expression of

targeted endogenous genes can be controlled.[96,97,301] Thus, CRISPR/Cas9 can be used to investigate phenotypes associated with gene activation as well as silencing (Fig. 6.2). In comparison to KO screens, so-called CRISPR activation (CRISPRa) is still in its infancy, with few published *in vivo* studies.[302-304] When combined with an *in vivo* experimental model, CRISPRa technology offers a novel approach to screen for regulators of metastasis.

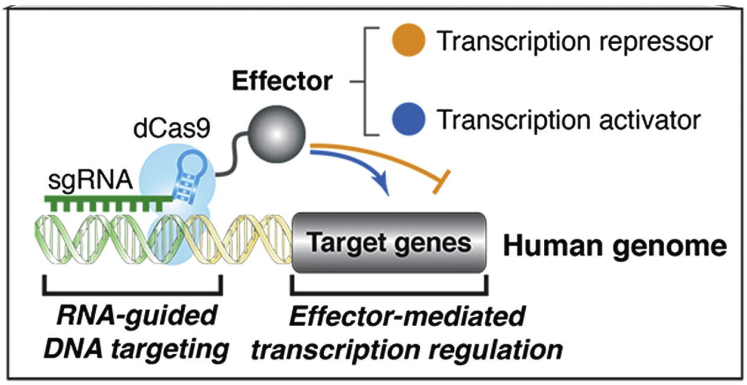


**Figure 6.2. CRISPR-mediated transcriptional regulation.** Catalytically inactive dCas9 can be fused to an effector domain and when bound to a gRNA, the protein is directed to a specific genomic region. Depending on which effector domain is used, transcriptional activators or repressors will be recruited, leading to regulation in expression of an endogenous gene. [Figure adapted from Qi et al.] [96]

### 6.1.5 Aims of this project

The genes and signalling pathways that must be initiated in order for a melanoma cell to survive and proliferate at a secondary site are not well understood. In this project, we aim to identify and characterise the genes that are upregulated in a melanoma cell in order for it to effectively colonise the lung. To do this we will:

- Use a CRISPRa library of gRNAs against mouse cell surface genes. We have chosen to specifically target cell surface proteins as they represent clinically tractable targets for immunotherapy, similar to the rationale for Trastuzumab (Herceptin), in which the monoclonal antibody binds to the HER2 receptors on breast cancer cells and both inhibits the normal function of the HER2 receptor and induces immune-mediated killing.[10]
- Use the B16-F0 mouse melanoma cell line. We chose this cell line as it is weakly metastatic, thus only having a low level of ability to effectively colonise the lung after tail vein dosing, and will allow us to identify genes that can enhance this ability.
- Perform a CRISPRa screen *in vivo* using the experimental metastasis assay. We chose this assay as using an *in vivo* protocol will ensure key regulators of metastatic colonisation are

present (the immune system, stroma, etc.) and the experimental metastasis assay avoids the gRNAs also having an effect on primary tumour growth (which would have been an issue with the spontaneous metastasis assay).

- Develop a suitable analysis pipeline for *in vivo* CRISPRa screen data.
- Validate candidate enhancers of metastatic colonisation identified by the screen.

This project was led by Dr Louise van der Weyden (LvdW), a Senior Staff Scientist in the Adams' lab. I carried out all lentiviral production, transductions, cell culture and flow cytometry during the *in vitro* phase of the screen and subsequent validation experiments. All other work was performed by LvdW, who has kindly allowed me to include this unpublished information.

# 6.2 Results

## 6.2.1 Preparation for genome-wide CRISPRa screen

The Weissman lab designed a genome-scale gRNA activation library targeting the mouse protein-coding transcriptome, designated mCRISPRa-v2.[305] This was designed using a comprehensive model to predict gRNAs with high activity, based on features such as nucleosome occupancy and TSS position. They also established subpools of this CRISPRa library which target specific functional gene groups. As we were interested in cell surface genes, for this project we used the 'membrane proteins (m6)' subpooled library; this consists of 10,975 gRNAs targeting 2,104 genes that encode for membrane proteins, and 250 non-targeting control gRNAs.

The m6 library was acquired from Addgene and expanded as described in Section 7.17.2 (backbone shown in Fig. 6.3). Plasmid DNA was extracted and packaged into a lentiviral vector. LvdW transfected B16-F0 mouse melanoma cells with an activator construct that contained two VP64 activation domains fused to dCas9 (Addgene #113341).[306] Blasticidin was used to select for and maintain the dCas9-expressing population (termed 'B16-F0-dCas9' cells). A titration was performed by transducing B16-F0-dCas9 cells with various volumes of the m6 library lentivirus (as described in Section 7.17.3). After 48 hours, BFP expression was measured by flow cytometry. The aim was to achieve an MOI of 0.2-0.5, to ensure that each cell carried only a single gRNA. The volume of lentivirus required to obtain an MOI of 0.3 (30% BFP-positive cells) was calculated and scaled up for the screen.

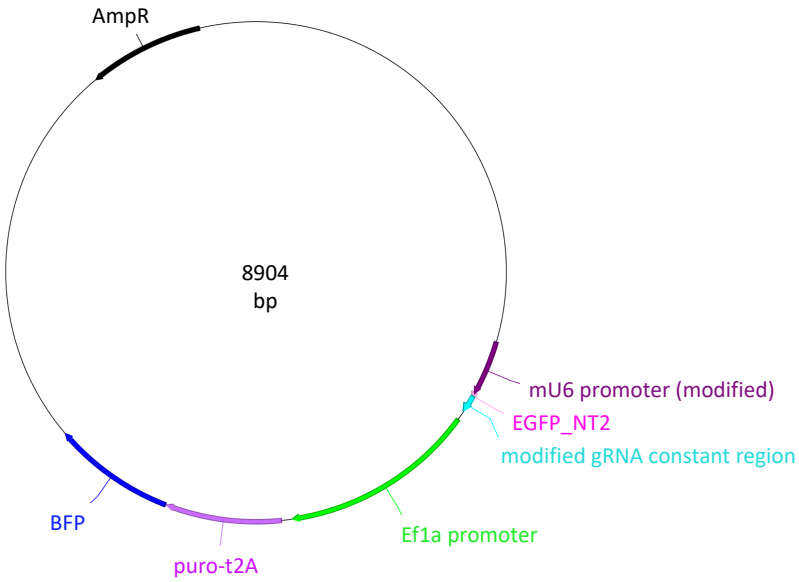


**Figure 6.3. pCRISPRia-v2 plasmid map.** The key features of the Weissman lab library backbone plasmid are shown. gRNAs are under the control of a mouse U6 (mU6) promoter. An Ef1a promoter drives expression of puromycin resistance (puro-t2A) and BFP.

### 6.2.2 Screening in an experimental metastasis model

The highest library coverage we could achieve per mouse was 50x ($5.5x10^5$ cells), as tail vein administration of a higher amount of this cell line could result in a tumour embolism that would be lethal to the mouse. We aimed for a higher *in vitro* coverage to minimise loss of coverage prior to moving *in vivo*. The initial target was ~300x and so $12x10^6$ cells were transduced at an MOI of 0.3. Details of the full screening protocol are provided in Section 7.17.4. After 48 hours, cells were passaged and puromycin was added to select for successfully transduced cells. A small sample was taken for flow cytometry; a transduction rate of 20% was achieved i.e. 200x library coverage (Fig. 6.4). Although lower than expected, this was still largely in excess of the required 50x and so we proceeded with the screen. Cells were passaged after 4 days in culture with puromycin and flow cytometry confirmed that selection was complete, with an almost 100% BFP-positive population (Fig. 6.4). Cells were collected for dosing at 9 days post-transduction, allowing enough time for the lentivirus particles to be cleared and for gene activation to occur.

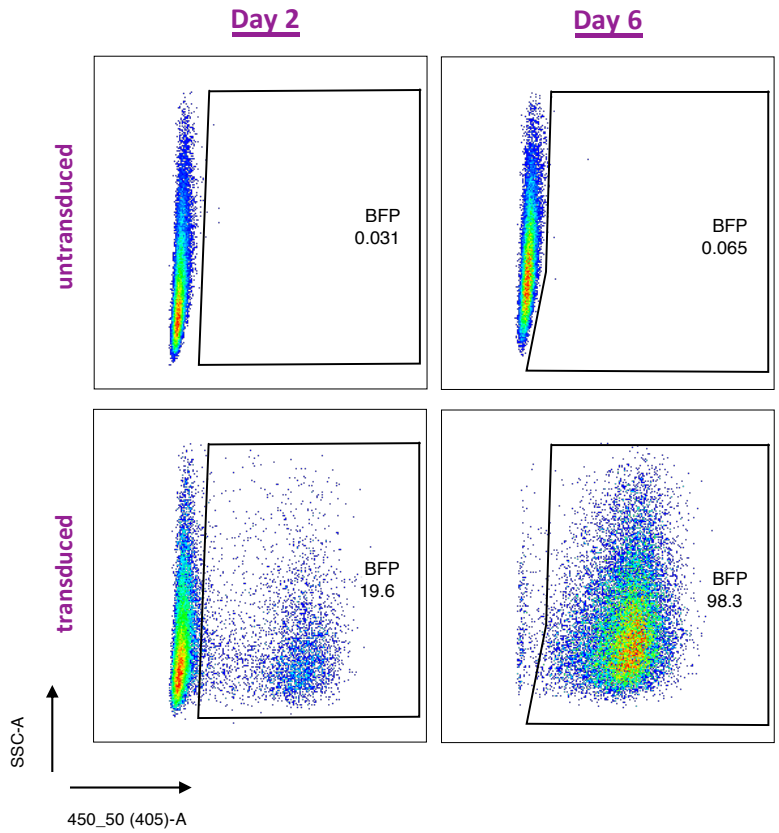


**Figure 6.4. Expression of CRISPRa library.** B16-F0-dCas9 cells were transduced with the m6 library lentivirus and analysed by flow cytometry on day 2 post-transduction. Puromycin was added and cells were analysed again on day 6 post-transduction. BFP expression (450_50 (405)-A) was measured, using untransduced cells as a control to gate the positive population.

After dilution in phosphate-buffered saline (PBS), $5.5x10^5$ cells were intravenously administered into the tail veins of 70 mice. As a control, some cells were also pelleted and frozen so that the baseline gRNA abundance could be measured. Activation of some genes may have altered cell proliferation/survival *in vitro*, therefore it was important that these effects were accounted for and did not interfere with the signal from gRNAs that impacted metastatic ability *in vivo*. Lungs were collected from mice at 2 timepoints: 35 mice were collected at 4 hours post-dosing (to determine what proportion of gRNAs successfully made it into the lungs) and 35 mice were collected at day 19 post-dosing (by this point culling was necessary as mice began to show symptoms associated with pulmonary tumour burden, such as rapid breathing). The entire lungs were collected and homogenised, and a portion of the homogenate taken for genomic DNA (gDNA) extraction. PCR was performed on the gDNA to amplify the gRNAs present in each sample (lung). PCR was also performed on gDNA from the cells that were collected on the day of dosing, and on the library plasmid DNA. The gRNAs present in each sample were identified by sequencing on a HiSeq 2500.

### 6.2.3 Analysis of CRISPRa screen data

As there is no well-established method for analysing *in vivo* CRISPRa screens, Vivek Iyer (a senior bioinformatician in the Adams' lab) trialled three independent analysis strategies to identify potential hits. The results from these methods were variable with little overlap overall. Design of gRNAs for CRISPR activation is challenging as transcription start sites are not well annotated for many genes. A pilot experiment performed by LvdW demonstrated that only 1 in 5 of the gRNAs in the m6 library that targeted the *CD8a* gene caused up-regulation of the CD8a protein (Table 6.1). Considering this, combined with the additional noise of doing a low coverage screen in a difficult setting (*in vivo*), two of the three bioinformatic analyses focused on the individual gRNAs rather than combining gRNAs to get a 'per gene' result.

**Table 6.1. Up-regulation of *CD8a* using m6 library gRNAs.** B16-F0-dCas9 cells were transfected with gRNAs targeting *CD8a* or non-targeting controls (NT_1-3). Expression of BFP and CD8a were measured by flow cytometry, using untransfected cells as a control for gating. BFP was used as a measure of the transfection rate, acting as a proxy for gRNA expression. The % of cells that were positive for each protein is shown. This experiment was performed by LvdW.

| gRNA | % BFP+ | % CD8a+ |
|---|---|---|
| untransfected | 0.11 | 0 |
| NT_1 | 77.4 | 0.13 |
| NT_2 | 31.6 | 0 |
| NT_3 | 86.1 | 0.14 |
| CD8a_1 | 86.4 | 63.8 |
| CD8a_2 | 71.2 | 0.2 |
| CD8a_3 | 78.6 | 0.19 |
| CD8a_4 | 75.8 | 0.51 |
| CD8a_5 | 51.2 | 0.7 |

#### 6.2.3.1 '98$^{th}$ percentile' method

For each day 19 mouse sample, gRNAs were ranked by their relative abundance (based on read count) and those in the 98$^{th}$ percentile (top 2%) were recorded. Those gRNAs that ranked highly in multiple mice were of most interest. Any gRNA which was present in < 50% of the day 19 samples, regardless of ranking, was disregarded. Additionally, a Z-score was calculated for each gRNA, comparing read counts from day 19 to the cells collected prior to dosing (day 0):

$$\text{Z-score} = \frac{\text{(average read count at day 19) - (average read count at day 0)}}{\text{Standard Error of the Mean for day 19 samples}}$$

To identify potential hits, gRNAs were first ranked by the number of mice in which they were present in the 98$^{th}$ percentile, and then by their Z-score (Table 6.2).

**Table 6.2 Screen analysis: '98$^{th}$ percentile' method.** Top 10 hits from analysis using the 98$^{th}$ percentile method.

| gRNA | Present in 98% percentile (no. of mice) | Present in day 19 sample (no. of mice) | Z-score |
|---|---|---|---|
| Lrrn4cl_+_8850757.23 | 4 | 17 | 3.19 |
| Slc4a3_+_75546398.23 | 3 | 20 | 3.92 |
| Tmem194b_+_52630698.23 | 2 | 23 | 2.97 |
| Zmynd12_+_119422735.23 | 2 | 19 | 2.09 |
| Fut2_-_45666416.23 | 2 | 14 | 2.09 |
| Smco3_+_136835494.23 | 2 | 22 | 2.05 |
| Tango6_+_106683055.23 | 2 | 14 | 2.01 |
| Gpr27_+_99692177.23 | 2 | 26 | 1.78 |
| Olfr1360_+_21675377.23 | 2 | 18 | 1.67 |
| Gpr75_-_30885367.23 | 2 | 21 | 1.67 |

### 6.2.3.2 JACKS method

A second analysis using the JACKS package[112] produced a different list of potential hits, with results obtained for genes rather than single gRNAs. A Z-score was calculated as described for the '98$^{th}$ percentile' method, but the average read count in the lung samples at 4 hours post-dosing was subtracted rather than using the read count in the pre-dosing cell samples from day 0 (Table 6.3).

**Table 6.3. Screen analysis: JACKS method.** Top 10 hits from analysis using the JACKS method.

| Gene | Z-score |
|---|---|
| *Lrrn4cl* | 0.76 |
| *Slc22a30* | 0.21 |
| *Tm4sf19* | 0.60 |
| *Kcnd1* | 0.77 |
| *Olfr323* | 0.76 |
| *Tmem54* | 0.48 |
| *Kcna5* | 0.47 |
| *Rhbdd2* | 0.39 |
| *Rxfp2* | 0.49 |
| *Scn8a* | 0.43 |

#### 6.2.3.3 'Appearance vs enrichment' method

A third analysis developed by Felicity Allen (a colleague in the Parts' lab, WSI) produced another list of potential hits. This analysis independently considered two factors: 'appearance' and 'enrichment'. 'Appearance' describes how many lung samples at day 19 carried a given gRNA; the presence of NTC gRNAs was used as a control to determine whether a targeting gRNA appeared significantly more than expected by chance. 'Enrichment' describes the abundance of a gRNA in a lung sample at day 19 relative to the abundance in the cells prior to administration to the mice on day 0. An 'enrichment' score was calculated by considering the difference in abundance and the variability of the data for this gRNA across all samples. For 'enrichment', the p-values for each gRNA were combined across all samples that the gRNA was present in. If a gRNA was significantly abundant in the lungs, or was simply present in a significant number of lungs, this could suggest that activation of that gene was advantageous for metastatic colonisation. The ranking of gRNAs varied depending on the weight given to each factor. Some gRNAs had high 'enrichment' but low 'appearance', and vice versa (Fig. 6.5). Others had moderately high scores for both factors. As this was a novel analysis strategy, we were unsure which ranking method was appropriate so we considered all possibilities.

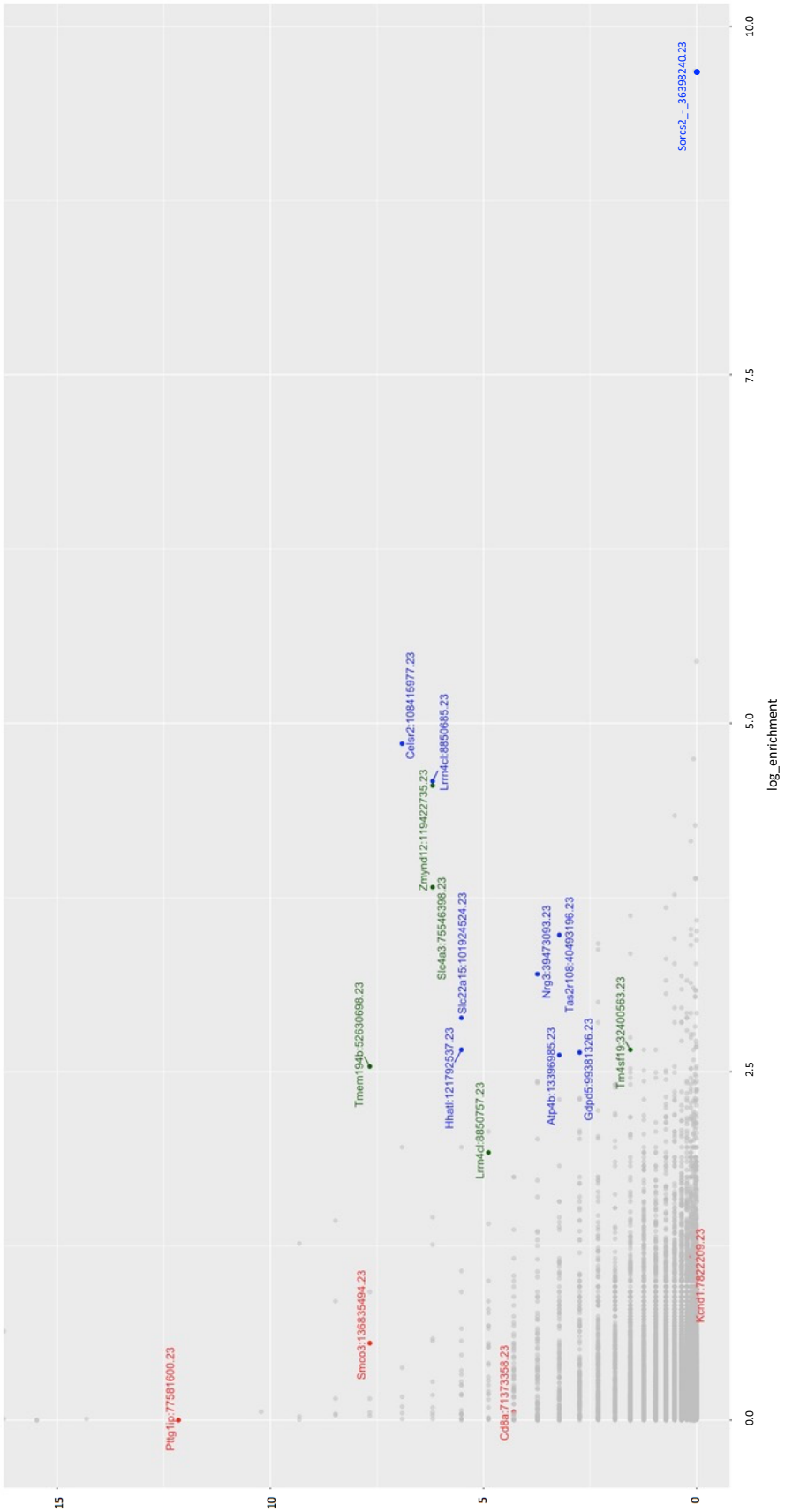


**Figure 6.5. Screen analysis: 'appearance vs enrichment' method.** Log-transformed 'enrichment' and 'appearance' scores are plotted for all gRNAs in the library. Several gRNAs of interest are highlighted and annotated.

### 6.2.4 Potential regulators of metastasis

We selected candidate genes from each of the three analyses for further validation (Table 6.4). From the '98$^{th}$ percentile' analysis, we chose to validate the top four scoring gRNAs which targeted *Lrrn4cl, Slc4a3, Tmem194b* and *Zmynd12. Lrrn4cl* was also the top hit in the JACKS analysis, and from this we additionally selected the third and fourth ranked genes, *Tm4sf19* and *Kcnd1*. The second ranked gene was *Slc22a30,* which appears to be a mouse-specific gene that arose from a recent duplication event[307]; we decided not to pursue this due to lack of human-relevance. From the third analysis we chose to validate gRNAs targeting *Pttg1ip* (high appearance), *Sorcs2* (high enrichment), *Celsr2* and *Lrrn4cl* (highest combined appearance and enrichment). The *Lrrn4cl* gRNA was different to the one identified by the other analyses. The *Slc4a3, Tmem194b* and *Zmynd12* gRNAs selected from the first analysis also had moderately high scores for both appearance and enrichment.

**Table 6.4. Candidates selected for validation.** The gRNAs chosen for validation are listed and the analysis method used to select each of these is indicated.

| Gene | gRNA ID | Analysis method |
|---|---|---|
| *Lrrn4cl* | Lrrn4cl_+_8850757.23-P1P2 | 98$^{th}$ percentile |
| *Slc4a3* | Slc4a3_+_75546398.23-P1P2 | 98$^{th}$ percentile |
| *Tmem194b* | Tmem194b_+_52630698.23-P1P2 | 98$^{th}$ percentile |
| *Zmynd12* | Zmynd12_+_119422735.23-P1P2 | 98$^{th}$ percentile |
| *Tm4sf19* | Tm4sf19_-_32400563.23-P1P2 | JACKS |
| *Kcnd1* | Kcnd1_-_7822209.23-P1P2 | JACKS |
| *Pttg1ip* | Pttg1ip_+_77581600.23-P1P2 | Appearance vs enrichment |
| *Sorcs2* | Sorcs2_-_36398240.23-P1P2 | Appearance vs enrichment |
| *Celsr2* | Celsr2_-_108415977.23-P1P2 | Appearance vs enrichment |
| *Lrrn4cl* | Lrrn4cl_+_8850685.23-P1P2 | Appearance vs enrichment |

#### 6.2.4.1 Biology of candidate genes

All of the genes selected for validation have a human homologue but the functions of many of the proteins are not well understood. Leucine Rich Repeat Neuronal 4 C-terminal-like (LRRN4CL) is a single-pass type I membrane protein. *TMEM194B*, also known as *NEMP2*, encodes for nuclear envelope integral membrane protein 2. *ZMYND12* encodes for Zinc finger MYND domain-containing protein 12. *TM4SF19* encodes for Transmembrane 4 L Six Family Member 19. Currently there is no literature regarding the biological function of these four

proteins, but all had consistently low expression in RNA-seq analysis of 27 normal human tissues.[308] Solute Carrier Family 4 Anion Exchanger Member 3 (SLC4A3) is a plasma membrane protein involved in exchange of chloride and bicarbonate.[309] It is most abundantly expressed in the brain and heart, and has been associated with idiopathic generalised epilepsy[310] and regulation of cardiac function.[311] Potassium Voltage-Gated Channel Subfamily D Member 1 (KCND1) is a multipass membrane protein. It is a pore-forming subunit of voltage-gated A-type potassium channels, which create currents in neurons and cardiac myocytes.[312] PTTG1IP is a single-pass type I membrane protein that binds to pituitary tumour-transforming 1 protein (PTTG1) and promotes its translocation to the nucleus.[313] It has been identified as an oncogene in endocrine cancers such as thyroid cancer,[298] and overexpression has also been shown in cancers such as breast[314] and colorectal.[300] It is has been to shown to decrease p53 stability by ubiquitination.[315,316] One study demonstrated that PTTG1IP interacts with CTTN to induce cellular invasion and migration in thyroid and breast cancer cells[317]. *SORCS2* encodes for vacuolar protein sorting 10 domain-containing receptor protein. It has roles in neuronal function and viability,[318,319] and has been linked to diseases such as bipolar disorder[320] and schizophrenia.[321] Cadherin EGF LAG seven-pass G-type receptor 2 (CELSR2) has been shown to have roles in ciliogenesis and neuronal development, including neuron migration and axon guidance.[322] Mutations in *CELSR2* have been associated with poor prognosis in endometrial cancer but there is no functional evidence for this.[323] *PTTG1IP* is the only candidate that has a clear functional association with metastasis, but several of the genes are not well characterised and may have interesting functions that have not been identified yet.

#### 6.2.4.2 Validation of candidate genes

LvdW cloned a gRNA targeting each candidate gene into the library backbone and packaged into lentiviral vectors (as described in Section 7.17.7.1). Three random NTC gRNAs from the library were also cloned, pooled and packaged into a single lentiviral vector for use as a negative control. All lentiviruses were titred in B16-F0-dCas9 cells (as described in Section 7.17.3). Each gRNA was transduced independently and cells were processed as in the screen: BFP was measured and puromycin was added after 48 hours, BFP was measured again on day 6 to confirm complete selection and on day 10 cells were collected for intravenous administration (as described in Section 7.17.7.2). Each mouse was dosed with $2x10^5$ cells, with 9-10 mice/condition. Lungs were collected after 10 days and melanocytic metastases were

counted by eye (as described in Section 7.17.7.3). Validation experiments were performed in batches with different mice cohorts, but the NTC was repeated in every batch.

Mice dosed with cells carrying the *Lrrn4cl, Slc4a3, Tmem194b* and *Zmynd12* gRNAs developed significantly more pulmonary metastases compared to the NTC (Fig. 6.6a). Similarly, cells expressing the *Tm4sf19* gRNA formed significantly higher numbers of metastases compared; the *Kcnd1* gRNA had a significant effect in one cohort ($p < 0.0001$), but not in another ($p = 0.5401$) (Fig. 6.6b&c). Mice dosed with cells carrying the second gRNA targeting *Lrrn4cl* or the *Sorcs2* gRNA had statistically significantly more metastases than the NTC, but a gRNA targeting *Celsr2* had no significant effect (Fig. 6.6d). Finally, there was no significant difference in the number of metastases formed by cells carrying a *Pttg1ip* gRNA compared to the NTC (the first *Lrrn4cl* gRNA was repeated in parallel as a positive control, Fig. 6.6e).

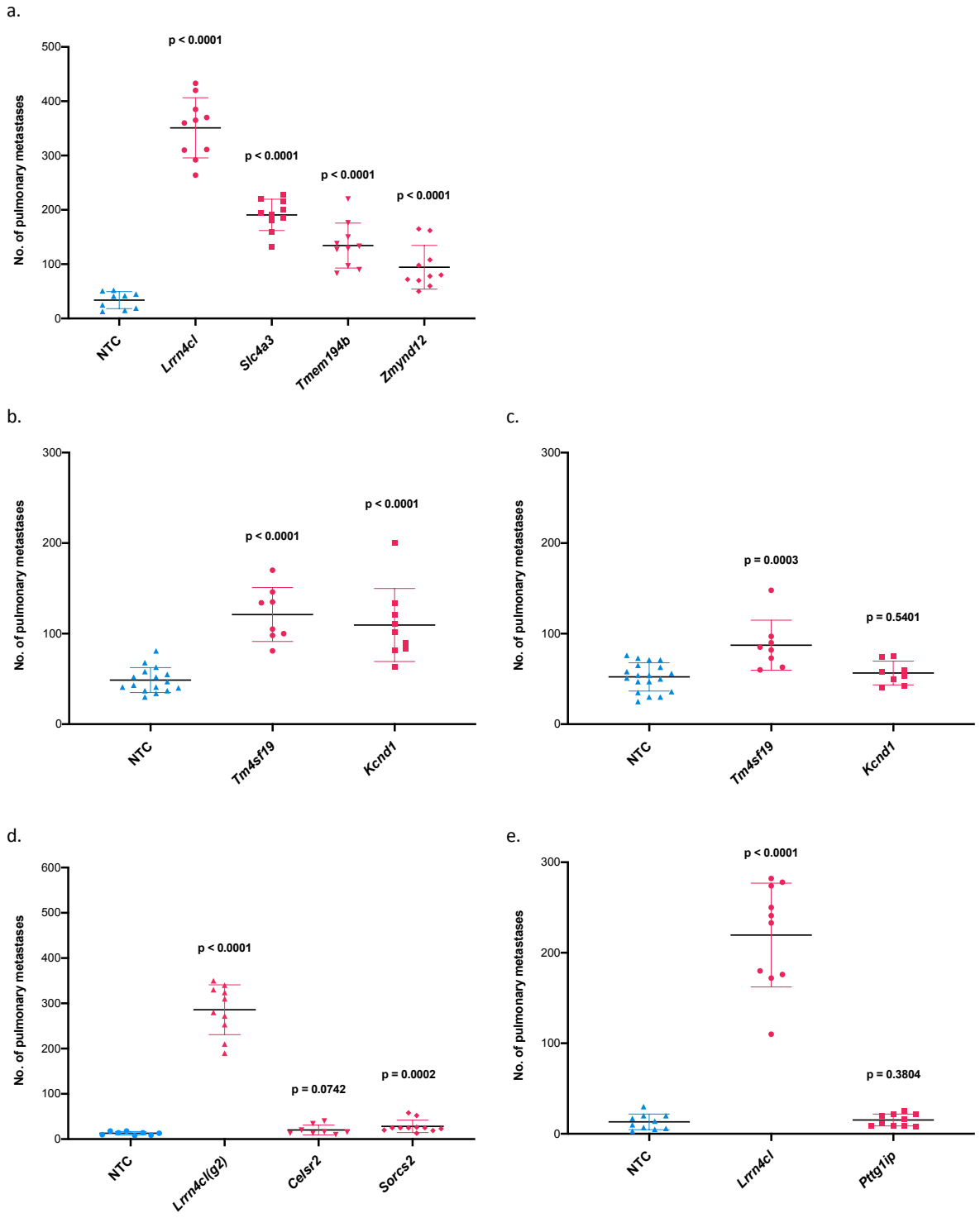


**Figure 6.6. *In vivo* validation of top scoring gRNAs from CRISPRa screen.** B16-F0-dCas9 cells were transduced with lentiviruses containing single targeting gRNAs, or a pool of NTC gRNAs. Transduced cells were intravenously administered to at least 8 mice per condition. On day 10 post-dosing, lungs were collected from the mice and metastases were counted by eye. The number of pulmonary metastases present in mice dosed with each gRNA are plotted; the mean is indicated by a line and error bars show the standard deviation. Each graph represents a single cohort. A Mann-Whitney test was performed to compare each gRNA with the NTC (p-values shown). *Lrrn4cl* and *Tmem194b* validated in 2 additional cohorts. *Slc4a3* validated in one independent cohort. *Zymnd12* was repeated in a further 4 cohorts, and validated in 3 of these. Data is not shown for these extra cohorts.

## 6.3 Discussion and future directions

In this project, we established a protocol for carrying out a pooled CRISPRa screen in B16-F0 melanoma cells. We extended this to an *in vivo* setting using an experimental metastasis model, allowing us to screen for genes that enhance metastatic pulmonary colonisation. The transfer from *in vitro* to *in vivo* was successful and using various analyses, we could identify gRNAs that were enriched in the lungs of mice. Despite little overlap in the results of the analysis strategies, several hits from all of them did validate in the experimental metastasis assay. As more data is produced, analysis of CRISPRa screens will undoubtedly become more accurate and refined but the approaches applied here were sufficient to identify true hits. Improvements to CRISPRa gRNA design should also increase the reliability of these data; variability in efficacy currently means that hits are often not supported by data from multiple gRNAs. Results from the 'appearance vs enrichment' method were the least robust, but it was unclear how best to interpret these data prior to validation. The strongest candidate gene was *Lrrn4cl*, which was a hit in all analyses; cells carrying gRNAs targeting this gene consistently formed more pulmonary metastases than any of the others.

Our data suggest that overexpression of *Lrrn4cl* enhances the ability of B16-F0 melanoma cells to colonise the lung. LvdW repeated the experimental metastasis assay using a *Lrrn4cl* cDNA (and empty vector as a control) as an alternative over-expression strategy and the result was replicated. This assay was also performed in immunodeficient mice lacking an adaptive immune system (B cells, T cells, NK cells) and a significant phenotype was still observed. LvdW repeated the cDNA assay in three additional melanoma cell lines, plus a breast, a colorectal and a bladder cancer cell line. Pulmonary metastases were increased in all lines, suggesting that the effect of *Lrrn4cl* activation is not specific to B16-F0 cells. The *Lrrn4cl* gene has a human ortholog with the same name (*LRRN4CL)* which encodes for a single-pass type I membrane protein. LvdW confirmed that this effect was not mouse-specific: enhanced pulmonary colonisation was observed in immunodeficient mice dosed with two human melanoma cell lines expressing a human *LRRN4CL* cDNA.

These findings suggest that LRRN4CL could be a potential target to reduce the development of pulmonary metastases in human cancer. RNA expression data from primary tumours indicates that *LRRN4CL* is highly enriched in melanoma (Fig. 6.7). Thus, it could have clinical relevance for the treatment of melanoma patients. Its location on the cell membrane also makes it an ideal drug target. To explore this, LvdW and Victoria Harle (a postdoc in the

Adams' lab) are investigating whether knocking out *LRRN4CL* in human melanoma cell lines can reduce pulmonary colonisation in mice.

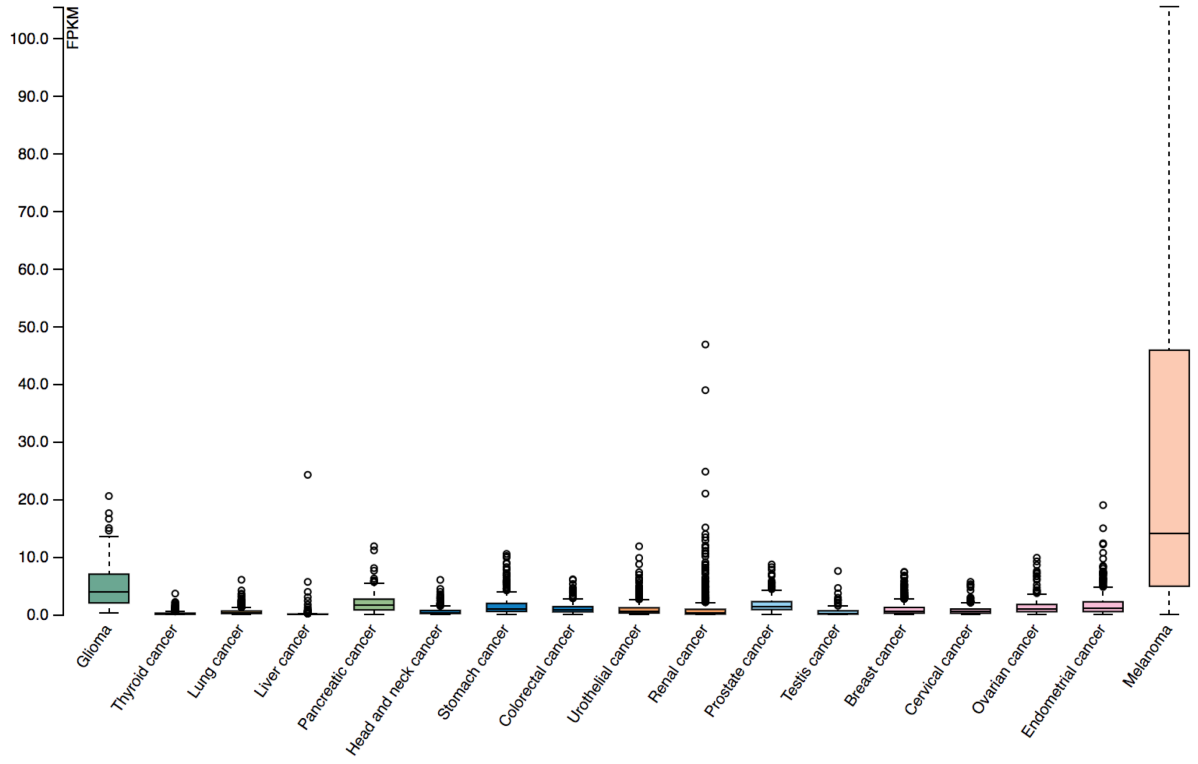


**Figure 6.7. Expression of *LRRN4CL* RNA in cancer.** RNA-seq data from 17 cancer types, reported as the median FPKM (number of Fragments Per Kilobase of exon per Million reads), was generated by TCGA. Graph taken from the Human Protein Atlas (available at proteinatlas.org/ENSG00000177363-LRRN4CL/pathology).

Limited information is available regarding this gene/protein in both mouse and human. In normal human tissues, the gene appears to be expressed at a low level (Fig. 6.8). As its normal cellular function is unclear, it is difficult to infer the mechanism by which activation of this gene could cause increased metastatic colonisation. Elucidation of any proteins that interact with LRRN4CL could reveal a wider network of genes with a similar role in metastasis. LvdW observed the *Lrrn4cl* phenotype in immunodeficient mice suggesting that, at least in mouse, increased colonisation is not due to an interaction with the adaptive immune system. Further investigation is required to understand the role of LRRN4CL in regulating metastatic colonisation and its potential as a therapeutic target.

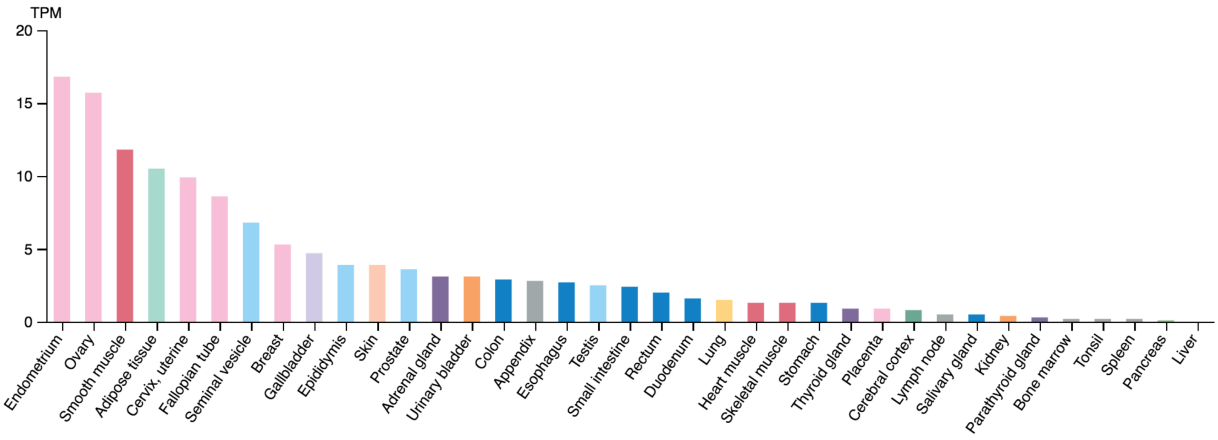


**Figure 6.8. Expression of *LRRN4CL* RNA in normal tissue.** HPA RNA-seq data from 172 normal tissue samples. Expression is reported as mean pTPM (protein-coding transcripts per million), corresponding to mean values of the individual samples for each tissue type. Graph taken from the Human Protein Atlas (available at https://www.proteinatlas.org/ENSG00000177363-LRRN4CL/tissue).

## 6.4 Final conclusion

CRISPR/Cas9 screening is proving to be an invaluable tool for functional genetic studies. Whilst its application in identifying fitness genes in cancer cell lines is fairly well-established, it has potential for many other uses that are still in their infancy. In this thesis, I have discussed two applications of CRISPR/Cas9 which have not been widely explored: knockout screening in iPSCs and activation screening in an *in vivo* model. Both of these proved to be technically challenging and produced lower quality data than more established strategies. We gained insights into the issues associated with screening in iPSCs and with further work, the data quality could be improved and these cells could be a useful model for studying genetic interactions. Analysis of *in vivo* screens and CRISPRa screens is not well-established but will undoubtedly become more refined as more data is produced. Despite this, our results were very promising and we successfully validated several hits that could further our understanding of metastatic colonisation. As the CRISPR toolkit improves and expands, so too will our ability to explore areas of cancer biology in a way that was not possible with other technologies.