Dissecting the Molecular Pathogenesis of Squamous Cell Lung Carcinoma

Can Altunbulakli

Asthma, Allergy & Lung Biology Research Division
School of Medicine, King's College London, UK

can.altunbulakli@kcl.ac.uk

Abstract. Squamous Cell Lung Carcinoma (SQC) constitutes 40% of UK cases of lung cancer. There are currently no specific therapies, valid therapeutic targets or methods for the early detection/chemoprevention of SQC. Amplification of *SOX2* and *FGFR1* have recently been implicated as driver events for SQC. I have used RNAi to understand the importance of these putative oncogenes in squamous carcinoma cell lines and have developed an assay that will help understand the molecular context in which oncogene activation may drive squamous carcinogenesis.
There were two main experimental strands:

1) I performed RNAi knockdown studies of *SOX2* and *FGFR1* on malignant cell lines with well characterized amplification of *SOX2* and *FGFR1*.In these cell lines; knockdown of SOX2 and FGFR1 inhibited cell proliferation. In addition SOX2/FGFR1 knockdown appeared to cause autophagy. I explored the potential for synergy of *SOX2/FGFR1* knockdown with an autophagy inhibitor in causing cell death.

2) Novel primer sets were designed and validated to analyse somatic mutations and copy-number variations in two key tumor suppressor genes, LKB1 and P16 were designed and validated. These primers were optimized for single molecule digital PCR to facilitate the analysis of these loci in archived formalin-fixed preinvasive lesions.

Keywords: Squamous Cell Lung Carcinoma, SOX2, FGFR1, LKB1, STK11, P16, digital PCR, Microdissection Molecular Copy Counting, RNAi screening

Word Count: 12485

1 Introduction

Lung cancer has long been the greatest cause of cancer related deaths worldwide (Landis et al. 1999). Human Lung Cancer comprises two main histopathological groups: Non-small cell lung cancers (NSCLC); which make up 80% of Lung Cancer cases, and small cell lung cancers (SCLC) which constitute the remaining 20% of total cases. NSCLC are subdivided into Adenocarcinomas (ADC), Large Cell Carcinomas (LCC), Bronchialveolar and Squamous Cell Carcinomas (SQC); SCLC cases generally show characteristics of neuroendocrine differentiation (Travis et al. 2002).

In recent years new targeted therapies have emerged for the treatment of NSCLC, especially with the identification of mutations in Epidermal Growth Factor Receptor (EGFR) in adenocarcinomas. Inhibitors of EGFR, prime among them gefitinib and erlotinib, are now used to inhibit the intracellular tyrosine kinase activity of the EGFR. EFGR inhibitors have shown demonstrable antitumor activity, in tumours with mutations or gross amplification of EGFR (Spiro et al. 2005). However tumours with EGFR mutation or amplification constitute less than 10% of total cases, particularly in the subdivision of ADC. In contrast, there are no targeted therapies available for the treatment of SQC and no adequate mouse models exist, although some are in development.

Distinct phenotypes of the NCSLC cancer types and their anatomical distribution suggest that the specific genetic alteration and the cell type in which the alterations occur together determine the tumor phenotype SQC tends to be a disease of the central airways and the most likely cell of origin is the keratin-5 expressing basal cell. Consistent with this, the morphology and histopathologic analysis of Mouse Lung SCC lesions reveals a striking resemblance of malignant cells to mouse tracheal basal cell progenitors (Sutherland et al. 2010).

It is postulated that SQC arises via a series of sequential histopathologic changes. The morphologic changes in the preneoplastic stage of SQC include hyperplasia, metaplasia of the squamous cells, dysplasia and carcinoma *in situ* (CIS) (Figure 1) Hyperplasia is indicated by the excessive accumulation of cells with normal phenotype in the tissue, and metaplasia can be observed by replacement of the bottom layer epithelium by squamous cells.

Normal architecture of the tissue changes in the dysplasia and carcinoma in situ (CIS) stages, during which the tissue is composed of cells of abnormal cells with large nuclei and prominent nucleoli. Along with these histologic changes, there is a progressive increase in molecular abnormalities corresponding to this multistage development of SQC (Wistuba et al. 1999).

The genetic events that are necessary for this histological progression and their timing relative to each other remain unclear. In this project I first explored the effects of RNAi-mediated knockdown of two putative oncogenes in a series of cell lines with variable amplification of two putative oncogenes, SOX2 and FGFR1. In parallel, I developed gene specific single molecule digital PCR assays that will allow the molecular context in which oncogene activation occurs to be elicited directly on heterogeneous formalin-fixed material.

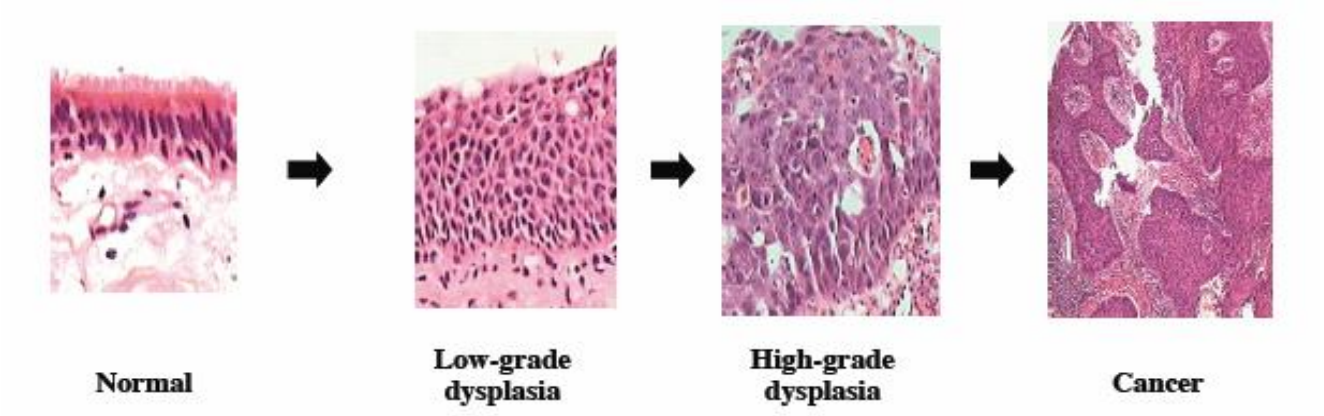


Fig. 1. Representative samples of SQC in different histologic phases. Histologically normal epithelium, Low Grade dysplasia, High Grade Dysplasia, Carcinoma *in situ* (CIS) (Adapted from McCaughan, Unpublished Data)

Potential Driver Oncogenes in SQC

A number of putative oncogenes have been implicated in SQC. However recent evidence has suggested that two of the most likely drivers are SOX2 and FGFR1. (Hussent et al. 2010, McCaughan et al. 2010)

The transcription factor SOX2, which is critical for the maintenance and differentiation of main airway basal cells; is grossly amplified in copy number in SQC lesions (Bass et al. 2009).

SOX2 (Sex Determining Region Y-Box 2) is a transcription factor composed of 317 amino acids and contains and HMG (High Mobility Group box) domain. It is an important transcription regulator of the stem cell phenotype in esophageal and tracheal stem cell populations (Que et al. 2007, Que et al .2009). It also controls stem cell phenotype in Neural Stem Cells (NSC) and is responsible for branching morphogenesis and epithelial cell differentiation in the developing lung (Hussenet et al 2010).

Fibroblast growth factors (FGFs) and their receptors (FGFRs) are crucial in mediating signals for the development, survival and proliferation of bronchial and tracheal epithelia. Oncogenic function of FGFRs has been implicated by the identification of copy number amplifications, chromosomal translocations and mutation accumulations along the FGFR gene locus in the cancer genome (Weiss et al. 2010). Amplification of the FGFR1 locus (8p11) is very common in epithelial carcinomas, particularly breast cancer and SQC (Turner et al. 2010).Importantly FGFR1 inhibitors are available and are currently being used in early clinical trials. The recent findings of the gross amplification of FGFR1 in 15-20% of lung SQC make FGFR1 dependent carcinoma cells prime candidates for therapeutic targets for FGFR inhibitors (Weiss et al. 2010) much like the case of EGFR amplifications for other subgroups of NSCLC.

In the light of the recent findings, I undertook a study examining the effect of RNAi mediated knockdown of SOX2 and FGFR1 on the morphology and phenotype of SQC cell lines. By using this approach in cell lines with amplification of these loci, the effect of depletion of SOX2 and FGFR1 can be interrogated(Weiss et al. 2010) Further, preliminary data in a single cell line suggested that SOX2 inhibition may induce autophagy. Autophagy inhibitors are available and have already been used in early therapeutic trials in some cancers.

Autophagy is an intracellular self-defense mechanism to prevent toxic accumulation of unused proteins and organelles in a cell. These residues are sequestered into autophagic vesicles and degraded via intracellular fusion with lysosomes. Autophagy is a particularly important survival mechanism for cancer cells, as they are faced with heightened cellular and metabolic stress. Thus autophagy, especially the inhibition of autophagy is considered an excellent therapeutic target in cancer. In healthy tissues; autophagy in basal levels contributes to the suppression of tumor development, as alteration of protein homeostasis, chronic tissue damage and organelle loss increase the production of reactive oxygen species (ROS) in a cell, contributing to DNA damage and accumulation of mutations which are the initiators of tumor formation and tissue dysplasia. In contrast, some malignant tissues autophagy clearly plays an important role in sustaining the survival of cancer cells (Amaravadi et al. 2011).

Potentially autophagy can sustain the elevated cellular metabolism of cancer cells via thecell viability via the recycling of intracellular components, particularly during nutrient limiting extracellular conditions (Rabinowitz et al. 2010).

In this project experiments were conducted to determine the combined effect of depletion of SOX2/FGFR1 and autophagy inhibitors, particularly chloroquine (CQ) on cell morphology and proliferation.The modulation of autophagy could potentially play a role in the therapy of SQC, perhaps in combination with SOX2 and FGFR1 depletion on SQC cell lines.

The antimalarial drug CQ is a potent autophagy inhibitor *in vivo* (Amaravadi et al. 2007). CQ and its derivatives are used for malaria treatment, rheumatoid arthritis (Kremer et al. 2001) and HIV (Romanelli et al. 2004). A drug capable of crossing the blood brain barrier, CQ can be safely escalated in dose in cancer patients, which makes it a prime choice for autophagy inhibition in cancer tissues (Gunja et al. 2009).

Digital PCR and understanding pathogenesis of squamous lung cancer

The genetic events that underpin the development of SQC are still not well understood. As stated above, recent evidence implicates SOX2 and FGFR1 in this process. However the molecular context in which this occurs is not understood. This information will be important in the development of rational preclinical models of SQC and in the development of biomarkers. To understand the molecular pathogenesis of SQC there is an overall aim in the laboratory to interrogate a unique archive of human preinvasive lesions (Figure 1) for a series of common somatic mutations and CNVs. These lesions are challenging as they are small, formalin-fixed and paraffin-embedded. Therefore a technique that has the capacity to interrogate small quantities of degraded DNA is required. I have used a novel single molecule digital PCR method that addresses these constraints, and the specific aim was to develop novel assays for two key tumor suppressor genes that may be critical in the development of SQC – LKB1 and P16.

PCR has the necessary sensitivity to detect and amplify a single copy of a target sequence if it is present in the reaction mixture. If we subject our template DNA to limiting dilution, so that the amount of template is less than a haploid genome per reaction, then the PCR results can be classified as positive (+) meaning the locus of interest is present in that reaction, or negative (-) meaning the absence of the marker. This binary readout is the origin of the name of the technique "digitalPCR": as the results can be explained in terms of "1"s and "0"s.Counting the number of positive reactions for a sequence of interest can be used to estimate the number of copies of that sequence in the template DNA used in the reaction (Daser et al.2006)

If a number of loci could be interrogated in parallel on the same DNA specimen that has undergone limiting dilution, then the relative copy number of one or more target loci of interest can be compared to the copy number of other loci with known copy number. This is achieved using a two phase PCR technique known as molecular copy counting (MCC). MCC is a direct derivation of digital PCR that allows us to define the relative copy number of multiple key loci in a single experiment simply by designing primer sets for those loci.

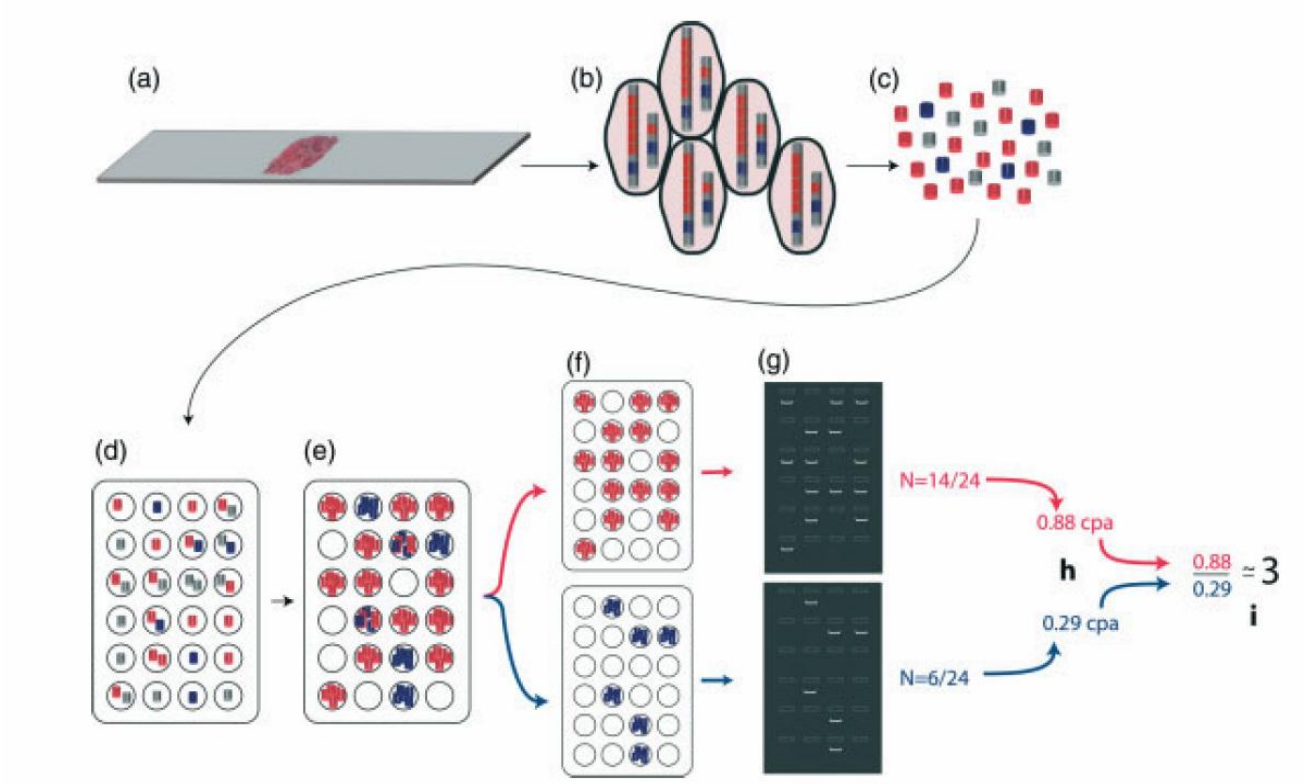


Fig. 2. Schematic representation of MCC using digital PCR and limiting dilution of template DNA to calculate copy-number. **(A-B)**Red marker indicates a locus of interest that has higher copy number relative to other loci; denoted in blue. **(C)** DNA is extracted into solution, and the red marker is in a higher concentration relative to the blue marker, which reflects their total copy number. **(D)** DNA solution then undergoes limiting dilution and distributed in even volume across a 96 well plate, so that there is less than 1 haploid genome per well. Red marker again reflects its original copy number relative to the blue marker, being present in more wells than the blue marker. **(E-F)** Each well is tested for the presence of red and blue markers in a two phase PCR. Multiplex first phase (e) allows for multiple markers to be assessed simultaneously, while the singleplex second phase amplifies each marker separately (f) **(G)** Each well for the marker of interest can give two results: positive or negative. Counting the positive results and comparing them to a marker of known copy number (blue) via Poisson distribution gives the relative copy number of the marker in question (see Methods). (Adapted from McCaughan et al. 2008).

Sequencing of PCR products is a standard technique. A potential advantage of digital PCR and the amplification of single molecules is that single DNA molecules can be sequenced and the relative frequency of mutant alleles in a DNA sample can be esti-mated. Further, as target sequences on single DNA molecules can be detected on a DNA template partitioned on multiple aliquots, even rare mutations will give positive results on a few reactions as opposed to PCR with bulk DNA, where the rare mutation will often be masked by the normal/unmutated sequence. Thus digital PCR gives us the ability to map copy gains, copy losses, mutation hotspots and frequencies for these events.

However the major source of DNA for the analysis for epithelial and soft tissue carcinomas is paraffin embedded biopsy samples. The tissue fixing process applied on the tissue samples is detrimental to the quality and quantity of DNA extracted from fixed lesions. . In order to understand the pathogenesis of SQC the earliest genetic events in preinvasive lesions need to be analyzed. The analysis of the tiny quantities of fragmented DNA available in heterogeneous preinvasive archived specimens is extremely challenging.

Degraded DNA used as a template would be predicted to result in longer amplimers being underrepresented, since they would suffer more from instances of DNA breakage within the target amplicon. Low quality and quantity of DNA eliminates the usefulness of bulk DNA PCR procedures. MCC technique used on DNA extracted from microdissected fixed tissue samples is termed microdissection MCC (µMCC). In µMCC primers used for the digital PCR process are designed to amplify shorter and more uniform (100-120bp) amplimer sequences (McCaughan et al. 2008). The technique has been successfully used to identify the minimum commonly amplified region (MCAR) in the recurrent 3q arm amplicon in high grade preinvasive bronchial lesions (McCaughan et al., 2010).µMCC studies also showed that dispersed lesions across the bronchial tree share amplicon boundaries and therefore have a clonal origin (McCaughan et al., 2011).

For this project I wanted to extend the number of key genes that could be interrogated in parallel on the limited biopsy material available using µMCC. In previous work developing a panel of SQC markers, primer sets for a number of loci including SOX2, FGFR1, KRAS, EGFR and TP53 were designed and used. However, other permissive genetic events that may allow oncogenes to drive the pathogenesis of SQC are not yet represented.

In particular I wanted to demonstrate the ability of µMCC technique to interrogate two known tumor suppressor genes: LKB1 (STK11) and P16 in preinvasive samples for the first time.

LKB1 (liver kinase B1, or STK11, serine/threonine kinase 11) is a potent tumor suppressor; with its germline mutation causing increased incidence of epithelial cancers, particularly colorectal (Peutz-Jeghers Syndrome) (Hearle at al. 2006). LKB1 inactivation is reported in many lung adenocarcinomas and SQC cases. In recent studies LKB1 deficient lung tumors were shown to have shorter latency, more frequent

metastasis, and accelerated tumorigenesis compared with the effects of other widely known tumor suppressor genes P53 and CDKN2A (P16) (Ji et al. 2007).

A recent study claims the chromosome loss for LKB/STK11 region in lung cancer approaching 100%, as the highest site of inactivation in lung cancer (Gill et al. 2011). 39% of primary NSCLC lesions have LKB1 inactivation by either homozygous deletion or point mutations in the gene, and more than 90% of lung cancer cases lose single or both copies in LKB1 gene locus (Gill et al. 2011).

CDN2KA is also a potent tumor suppressor gene that regulates the key pathways of TP53 and RB1.Shared coding regions and alternative reading frames allow CDNK2A to code for P16, a cyclin dependent kinase inhibitor, and p14 which has a role in the stabilization of p53 through the binding of MDM4 (Robertson et al. 1999). CDN2A resides in 9p21 (Nobori et al. 1994). The region of 9p21 is notably involved in chromosomal inversions, translocations, heterozygous deletions, and homozygous deletions in a variety of malignancies such as NSCLCs, melanomas and gliomas (Kamb et al. 1994), Furthermore; 50% of melanoma lesions investigated carried loss of 9p21 and the region was documented to undergo 100% loss of heterozygosity in SQC of the lung (Wiest et al. 1997).

2 Materials and Methods

2.1 RNAi Experiments

Cell Lines and Culture Conditions

Squamous Lung Cancer cell lines NCI H520, KYSE 140, LUDLU 1 and BEAS 2B were obtained from Dr. Frank McCaughan, KCL and Cambridge. NCI H520, KYSE 140 and LUDLU 1 cells were maintained in RPMI 1640 medium (Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies). BEAS 2B cells were maintained in Keratinocyte SFM medium (Life Technologies) supplemented with Bovine Pituitary Extract and Human Recombinant Epidermal Growth Factor (Life Technologies).

Table 1. Reported Mutations and Copy Number Changes in Cell lines cultured for experiments

Name	***Type***	***Genomic Aberrations***
NCI H520	Squamous Lung Carcinoma	deletions in *P16* G>A substitution in *TP53* Copy number amplification in *SOX2, FGFR1*
KYSE 140	Oesophageal Carcinoma	deletions in *P16* A>G substitution in *TP53* Copy number amplification in *SOX2*
LUDLU1	Squamous Lung Carcinoma	Normal copy number in *SOX2* *FGFR1* copy number unknown
BEAS 2B	Normal Bronchial Epithelium	Normal copy number in *SOX2, FGFR1*

RNAi Knockdown Assays

For RNA interference experiments, ON-TARGETplus SMARTpool small interfering RNAs (siRNAs) targeting *SOX2, FGFR1* and *GAPDH* along with a negative control (ON-TARGET plus siCONTROL nontargeting siRNA pool) were obtained from Dharmacon. Cells were seeded in 6-well/24-well culture plates at 1 x 10^4 cells/ml 24 h before transfection. Cells were transfected at a final concentration of 20 nM siRNA using Thermo Scientific DharmaFECT 4 transfection reagent according to the manufacturer's instructions. As previously been described, a dual knockdown strategy was employed (Schuck et al. 2004) so that 24 hours after the initial transfection, a second identical knockdown strategy was employed. The cells were then incubated at 37˚C for up to 96 hours before protein or cell proliferation assays. Morphology of cells was observed using a Zeiss AxioscopII type of light microscope and images were taken before cell lysis.

Western Blot Analysis of Protein Levels

Cells were washed once with cold 1x PBS (Life Technologies) and lysed with Cell Lysis Buffer NP-40 (Invitrogen) in the presence of protease inhibitor cocktail cOmplete Mini EDTA-free tablets (Roche). Protein concentrations of the cleared lysates were calculated via BCA protein assay kit (Pierce).

Cleared lysates were boiled at 95˚C for 5 min and volume containing 20µg/ml protein for each sample was electrophoresed via SDS-PAGE using 10% polyacrylamide gel and transferred to Nitrocellulose membrane (Bio-Rad) using Mini-PROTEAN tetra cell (Bio-Rad). Membranes pre-blocked with 5% non-fat milk in 1x TBS with 0.1% Tween 20 (TBS-T) were incubated with primary antibodies against *SOX2* (R&D Systems Biology, 1:1000 dilution), *FGFR1* (Santa Cruz Biotech, 1:1000 dilution)

αtubulin (R&D systems Biology (1:5000 dilution) or *LC3B (*Novus Biologicals, 1:1000 dilution).

After primary staining and 3 washes with TBS-T, membranes were incubated with horseradish-peroxidase conjugated rabbit anti-mouse (for anti *SOX2*, anti α-tubulin primary antibodies) or goat anti-rabbit (for anti *FGFR1* primary antibody) polyclonal secondary antibodies for 1 hour at room temperature (Dako, 1:1000 dilution).

After 3 washes with TBS-T, antibody binding on the membrane was visualized by electroluminescence kit (Amersham). Subsequently; bound antibodies were stripped from the membrane via buffer Re-blot Plus (Millipore) and re-stained with α-tubulin (R&D systems biology, 1:5000 dilution) monoclonal antibody as a loading control.

Chloroquine (CQ) treatment on siRNA knockdown cells

For Chloroquine (CQ) treatment on gene knockdown cells, Chloroquine Diphospahate salt was purchased from Sigma. CQ was diluted to 20 μM in RPMI-1640 medium supplemented with 10% FCS and added on cells plated on 96 well plates prepared for MTT proliferation assay, 200μl per well. Plates were incubated for 72 or 96 hours before MTT cell proliferation assay.

Cell Proliferation Assays

The 3-[4, 5-dimethylthiazol-2-yl]-2, 5-diphenyltetrazolium bromide (MTT) assay (Sigma) was used to determine the status of cell proliferation in siRNA and CQ treatment experiments. In both experiments, cells were diluted to a concentration of 1 x 10^4 cells/ml in RPMI 1640 medium supplemented with 10% FBS and seeded in triplicate onto 96 well plates, using 200μl of cell suspension per well. After siRNA knockdown and CQ treatment via the instructions above, the cells were then treated with 20 μl of 5 mg/ml MTT reagent in PBS for 4 h before adding 150μl of detergent reagent composed of 4Mm HCL, 0.1% NP-40 in Isopropanol to solubilize the formazan precipitate. The reaction product was then quantified by measuring absorbance at 570 nm with reference to 650 nm using an Anthos LabTec HTIII plate reader.

Statistical Analysis.

All statistical calculations for proliferation assays were done with the Students t-test using a $p>0.05$ threshold for significance.

2.2 Genome Interrogation via digitalPCR

Marker Identification

I constructed two different types of markers for the identification of genome aberrations in *LKB1* (*STK11*) and *P16* genes: Mutation Markers and Copy Number Markers.

Mutation Hotspots on the *LKB1* gene were identified via the COSMIC Database (Catalogue of Somatic Mutations in Cancer) of Wellcome Trust Sanger Institute (Figure 3). Mutation Markers are constructed as Primer sets surrounding mutation hotspots on the *LKB1* gene (Figure 5). When markers were optimized and selected; amplicons generated as the result of the PCR process were then sent to sequencing to discover mutated sequences on *LKB1* gene in SQC lesions. This was facilitated by the addition of generic M13 tags to one of the internal primer sequences.

Copy Number Markers were designed to amplify separate regions flanking the *LKB1* and *P16* genes. Unmasked Human *P16* and *LKB1* gene sequences and flanking sequences of 10kb length were imported from ENSEMBL genome database. Sequence files were opened with the Artemis Genome Browser and Annotation tool (Wellcome Trust Sanger Institute). Mutation Hotspots identified via COSMIC were manually annotated onto Gene Sequences on Artemis (Figure 5).

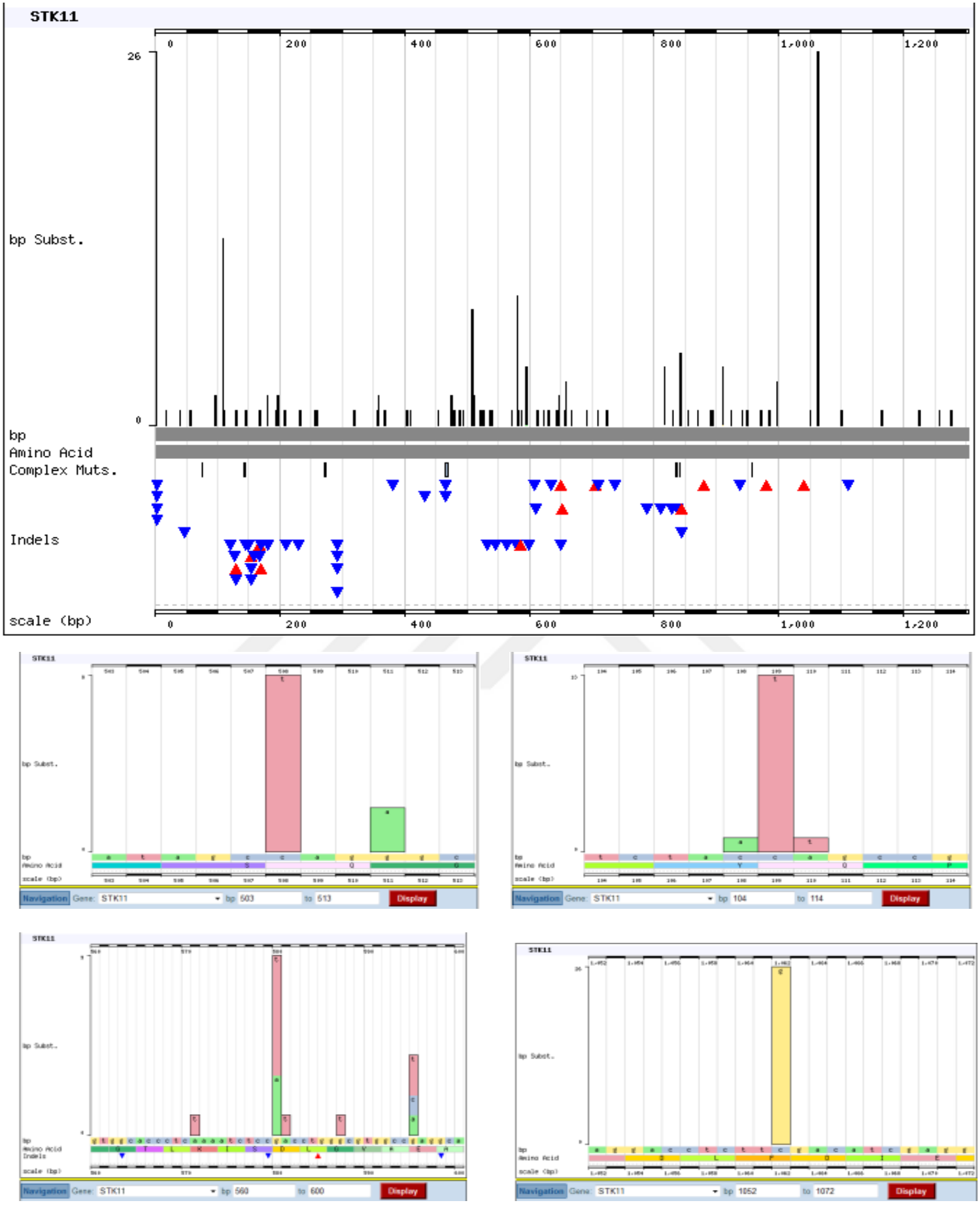


Fig .3. Mutation Hotspots identified on the *LKB1* (*STK11*) gene. **(A)** COSMIC database shows repeated mutations on 5 spots on the *LKB1* gene locus **(B)** c.109C>T 13 bp substitutions **(C)** C.508C>T 8bp substitutions **(D)** c.580G>T 9bp substitutions, c.595G>T 6 bp substitutions **(E)** c.1062C>G 26 bp substitutions.

Primer Design

Both classes of Primer Sets (Mutation Markers and Copy Number Markers) were generated using the Primer3 software webtool (Rozen et al. 2000). To generate a Mutation marker set, 200 base pairs (bps) of sequence surrounding the mutation hotspot was extracted in FASTA format from the Artemis Sequence file, and fed into Primer3. In Primer3; primer sequence selection was limited to the sequence flanking around the mutation hotspot, and primer sequences which included the mutation site were excluded to prevent primer overlaps with mutation hotspots.

Amplicon sequences were selected to be between 100-120 bps. A hybridization probe was selected along each primer set. Primer sequences were selected within a length of 18–22 bps, with a melting temperature (Tm) of 55–62˚C (calculations based on T_m = 2x (A + T) + 4x (G + C)). I aimed to use GC clamps of 1-3 bps at the 3' and 5' ends of each primer sequence, but eliminated GC clamp addition whenever other selection parameters proved to be of great constraint on primer generation.

Table .2. Parameters of contstraint for Primer Set generations

Parameters	***Values***
Minimum external amplicon length (bps)	100
Maximum external amplicom length (bps)	120
Minimum melting temperature (˚C)	55
Maximum melting Temperature (˚C)	62
5' GC clamp*	0-3
3' GC clamp*	0-3
Maximum polyX**	4
Minimum %GC	20
Maximum %GC	70

*GC clamp are the constituent bases of the 5' or 3' ends of the primer sequence that are G or C

** polyX refers to the maximum number of consecutive identical bases inside the primer sequence

Copy Number Primer sets were generated by feeding the entire gene sequence extracted from Artemis in FASTA format into Primer3 webtool. Amplicon length was again kept between 100-120 bps, and a hybridization oligo was generated along each primer set. Out of the oligonucleotide sets generated, 15 sets were selected for *LKB1* and *P16* genes. Sets were chosen to amplify evenly dispersed regions along each gene. For each gene, 2 sets of Copy Number primers were selected from the flanking sequences surrounding the gene.

Selected primer sets (Both Copy number and Mutation Markers) were then tested to exclude multi-copy sequences using the reverse electronic PCR (Reverse ePCR) web tool (National Centre for Biotechnology Information, NCBI) (Figure 4). Primer sets hitting more than one region in the genome, and sets allowing for 2 gaps and 2 mis-matches were eliminated. Copy number sets and mutation sets passing the hit test were manually annotating into the Artemis sequence file (Figure 5).

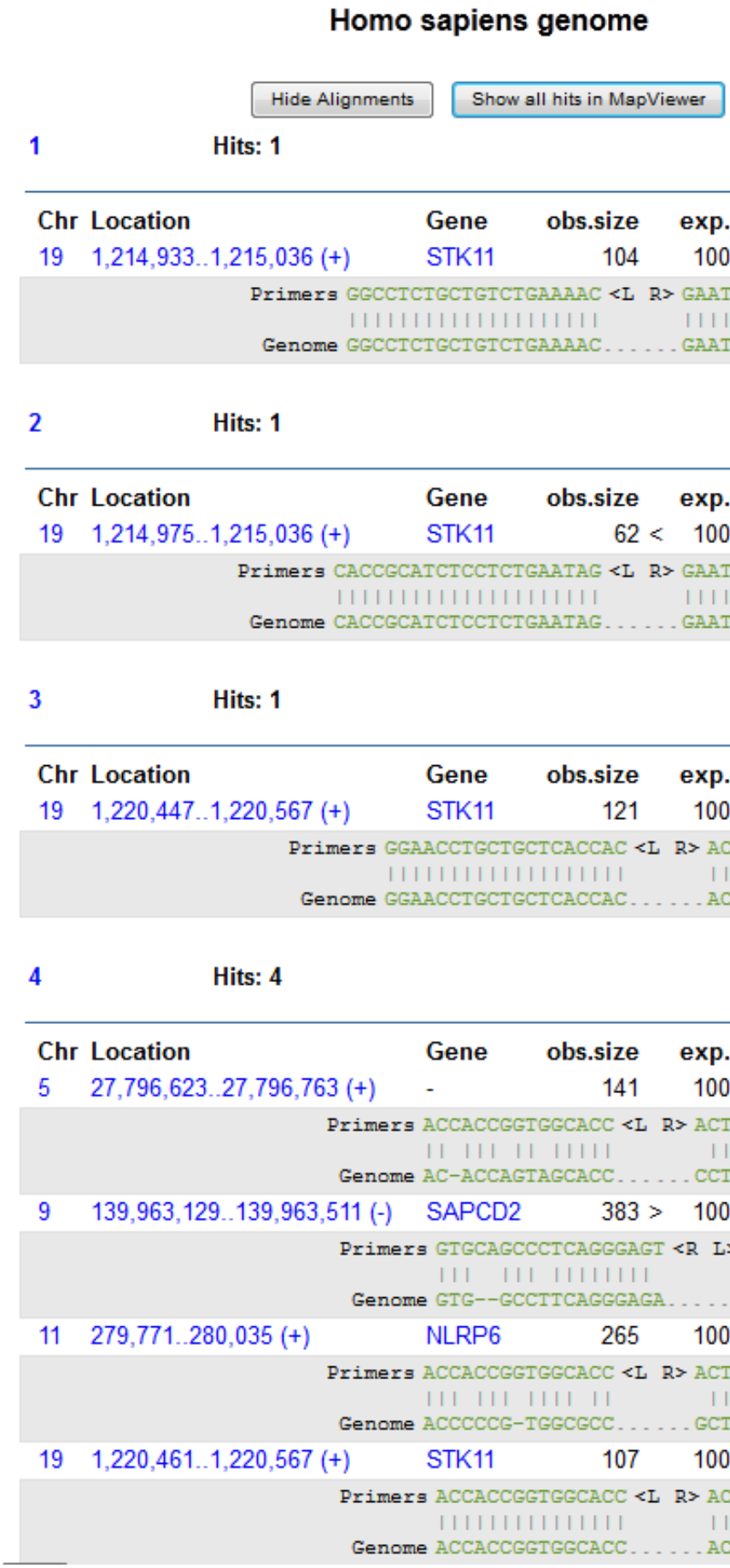


Fig.4. Sample Reverse ePCR of two primer sets: Copy number set (1-2) and a Mutation set (3-4). Forward-Reverse (1, 3) and Internal oligo-Reverse (2, 4) pairs were tested separately against multiple hits in the genome. Copy number set can be seen to hit only the *STK11* gene in Human genome, while the Mutation set hit 4 separate regions and was eliminated from the selected list of primers.

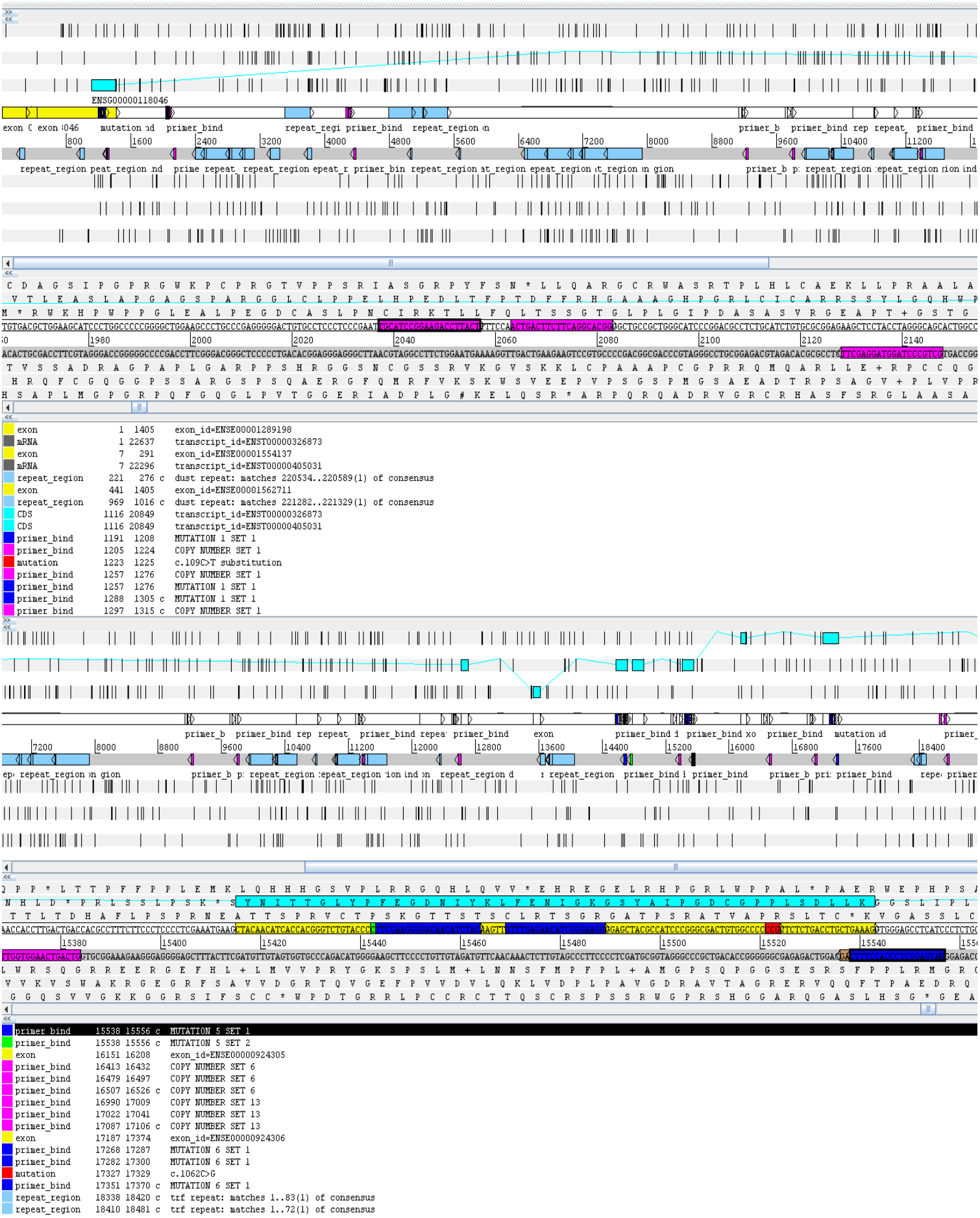


Fig.5. Sample of Annotated Primer Sequences (A Copy Number Set and a Mutation Marker set) on *LKB1* genomic sequence file. **(A)** Copy number set oligonucleotides (Pink) will amplify a region on the *LKB1* gene without any repeat sequences. **(B)** Mutation set oligonucleotides (Blue) surround the immediate vicinity of the mutation hotspot (Red) annotated on the sequences file.

Digital PCR assays

Selected Primer Sets as the result of the primer generation procedure were purchased from Sigma. DNA samples from patient biopsies were obtained from Dr. Frank McCaughan, and were extracted from paraffin embedded tissue samples according to methods in McCaughan et al. 2010.

For the first phase of the assay, all primer sets (forward-reverse pairs) were combined in a single multiplex PCR reaction. A master mix containing 1x Gold PCR Buffer, 2mM $MgCl_2$, 0.1 u/µl Taq Gold Polymerase, 200µM dNTP of each kind, combined with Forward and Reverse oligonucleotide pairs (0.15µM of each oligo) of each marker Sequence (Both Copy Number and Mutation Marker sets) was constructed. Into the master mix containing the primers, 0.03 genomes/µl concentration of genomic DNA was added. The final mix was distributed 10µl/well onto 88 wells of a 96 well plate. The remaining 8 wells received 10µl/well of the master mix without the DNA dilution, and were used as negative controls for the multiplex PCR reaction. The surface of each well of the 96 well plate was covered with a drop of heat conductive mineral oil to ensure efficient amplification.

Phase 1 PCR conditions that was used: 1cycle of 93˚C for 9 min, 25 cycles of 94˚C with 20sec/cycle, 1 cycle of 52˚C for 30sec and 1 min at 72˚C.

On the second phase, PCR products of the first phase were diluted to 500µl with Nuclease free water and 5µl of this DNA dilution was mixed into the newly made master mix as the DNA template. Components of the master mix were added as: 1x Gold PCR Buffer, 1.5mM $MgCl_2$, 0.1 u/µl Taq Gold Polymerase, 200µl dNTP of each kind, with 1µM oligonucleotide concentration (Internal oligo-Reverse pair). Separate mixes were prepared for each marker to be assayed.

As the second phase will assay each marker separately, mixes containing the primer sets were distributed onto 384 well plates, using customized robotics that were available offsite at the MRC in Cambridge. This equipment was used to prevent variations in volume, conserve time and because it was cost effective.

Thermocycling conditions for Phase 2 reactions: 1 cycle of 93˚C for 9 min, 33 cycles of 94˚C with 20sec/cycle, 1 cycle of 52˚C for 30sec and 1 min at 72˚C.

Phase 2 products were analyzed via Agarose Gel Electrophoresis, or 384 well plates undergoing melting curve analysis to identify positive wells for each primer set.

Melting curve analysis

In other experiments results were scored by melting-curve analysis. This was performed using the dissociation curve facility on a ABI Prism7900HT. The conditions were 500C for 2 minutes, then a temperature ramp between 65oC and 95oC using a 'ramp rate" of 2%. Custom-designed software (All-in-one MCA handler, KingCurve 2.0Blind, KingCurve7, Impose controls and Report npos – Paul Dear MRC Cambridge, unpublished) was used for data-handling and to call the melting-curves. KingCurve, the software designed for calling individual melting-curves (corresponding to individual wells) positive or negative had the facility to blind the user to the target sequence being called, thus reducing the possibility of any bias.

Copy number calculations

Raw MCC results are the total number of aliquots out of 88 (8 negative controls make up the 96 wells) that are positive for a specific target sequence. This number is converted into the number of copies per aliquot (cpa) using a calculation based on the Poisson distribution (Equation 1). For this calculation, it is assumed that the DNA molecules are distributed randomly among the 88 test aliquots. Further, although aliquots can only be positive or negative for the target sequence some positive aliquots may contain two or more copies. Therefore aliquots with 2 or more copies will score the same as those with a single copy. Using the Poisson distribution to allow for this, and calculating the probability that a well is negative for the target sequence of interest, the following equation can be derived (Equation 1).

$$1 = -\ln\left(1 - \frac{\mathrm{Np}}{\mathrm{N}}\right) \quad (1)$$

1: Number of copies per aliquot
$N_{P:}$ Number of positive aliquots
N: Number of total aliquots

Therefore, for each target sequence, the number of positive wells can be translated into the number of copies of that sequence per aliquot. In a single experiment in which multiple target sequences are tested the relative number of copies per aliquot will reflect the average relative copy-number of those sequences in the cells from which the DNA was extracted.

Sequencing

PCR products were scored as present or absent on the basis of the melting-curve analysis or gels. For the sequencing analysis, raw PCR products were chosen from wells known to have product on the basis of the melting-curve analysis and diluted 1:3 before being sent for sequencing using commercial providers (Source BioScience, Nottingham, UK). In a number of cases the internal primer was modified to have a M13F tag which allowed the short amplimers to be sequenced more readily.

3 Results

3.1 RNAi Experiments on SQC cell lines

KYSE 140 and NCI H520 cells were established in culture; based on having gross amplification of *SOX2* and *FGFR1* genes. KYSE 140 and NCI H520 cells are widely used lung cancer cell line with well documented characteristics of SQC, such as losses in tumour suppressor genes P53 and P16. My first aim was to establish a dose of siRNA that would have a high grade knockdown effect on *SOX2* and *FGFR1* on these SQC cell lines.

A range of siRNA concentrations between 10nM and 100nM were tested on KYSE 140 cells to establish optimal dose for *SOX2* and *FGFR1* knockdown. My aim was to avoid excess toxicity arising from high transfection reagent and siRNA concentration while achieving a high grade knockdown of my target genes. However because of issues of antibody specificity and difficulties in optimizing western blotting conditions, assessing the relative toxicity and efficiency of different concentrations of siRNA was limited to observations of transfected cells via light microscopy. Based on observations of cells after knockdown I decided on applying double knockdown; each with a concentration of 20nM. As KYSE and H520 cell lines possess high copy number of *SOX2* and *FGFR1*, it was decided that a double transfection with siRNA would be more efficient in knocking down target genes.

Early western blots resulted persistently in faint signals on film, when probing for FGFR1, SOX2 as well as the loading control antibody – αtubulin - stains, irrespective of the protein amount loaded onto SDS-PAGE, the antibody concentration used or the duration of exposure. To overcome this issue, I first adjusted the RNAi knockdown protocol by plating cells on 6 well plates, using a single well for each condition; to increase the number of cells and protein harvested. Previously; I was performing knockdown experiments on 96 well plates, plating cells in triplicate for each condition. While I had sufficient protein concentrations on all conditions to load 20µg/ml protein per SDS-PAGE well, this adjustment allowed me to achieve higher protein concentrations in cell lysates, but did not resolve my signal problem.

Antibody efficiency was tested by harvesting cells directly from culture flasks, acquiring lysates with a high amount of protein. Blotting and staining of these "bulk" lysates allowed me to discern clear bands using SOX2, FGFR1 and αtubulin antibodies after prolonged exposure times. However; considering the extremely high concentration of these bulk lysates (up to 2000μg/ml after 1:20 dilution, actual concentration likely much higher due to BCA standard curve saturation) and the overexpression of *SOX2* and *FGFR1* in KYSE and H520 cells, the issue of low signal clearly persisted.

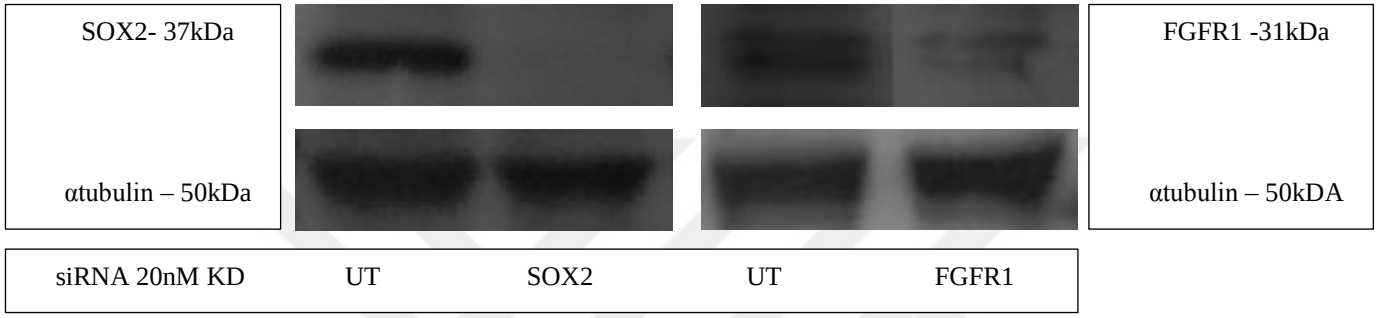


Fig .6. Successful SOX2 and FGFR1 knockdowns on KYSE 140 cells demonstrated with Western blot images **(A)** UT and SOX2 knockdown lysates stained for SOX2 **(B)** UT and FGFR1 knockdown lysates stained for FGFR1 **(C-D)** Loading controls of UT, SOX2 and FGFR1 knockdowns stained for αtubulin.

I solved this issue finally by changing the primary and secondary antibodies for all my target proteins, along with the ECL solution used thus far. I previously used a homemade ECL solution composed of 1.25mM Coumaric acid and 0.2 mM Luminol, and changed it to an ECL solution ordered from Amersham. Antibodies were kept at -20˚C freezer according to manufacturer's instructions and aliquoted to prevent them from degrading, but in retrospect it is likely that a known episode in which there was an unplanned and uncontrolled freezer defrost led to my antibodies degrading completely, causing the signal problem.

Despite losing substantial time, high grade knockdown of *SOX2* and *FGFR1* was soon proven in both KYSE 140 and NCI H520 cells (Figures 6, 7). Repeated experiments showed that double knockdown of 20nM siRNA was sufficient to effectively knockdown both *SOX2* and *FGFR1* on KYSE and H520 cell lines.

Having proven knockdown of my target genes, I then proceeded to determine the effect of depletion of these genes on cell proliferation by MTT assays. My aim was to find evidence of proliferation reduction on SQC cell via the inhibition of oncogenes with high copy number and high profile of expression. Reductions of up to 60% in proliferation ($p<0.05$) was achieved in H520 cells with *SOX2* knockdown in compari-

son to Untransfected (UT) cells; which were treated with only transfection reagent. Depletion of *FGFR1* resulted in a ~57% reduction in proliferation ($p<0.05$) compared to UT cells (Figure 8).

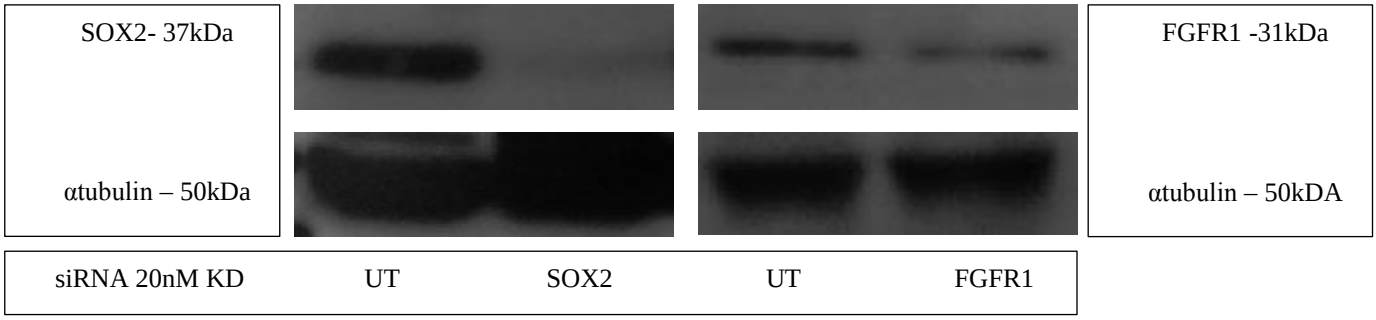


Fig.7. Successful SOX2 and FGFR1 knockdowns on NCI H520 cells demonstrated with Western blot images **(A)** UT and SOX2 knockdown lysates stained for SOX2 **(B)** UT and FGFR1 knockdown lysates stained for FGFR1 **(C-D)** Loading controls of UT, SOX2 and FGFR1 knockdowns stained for αtubulin.

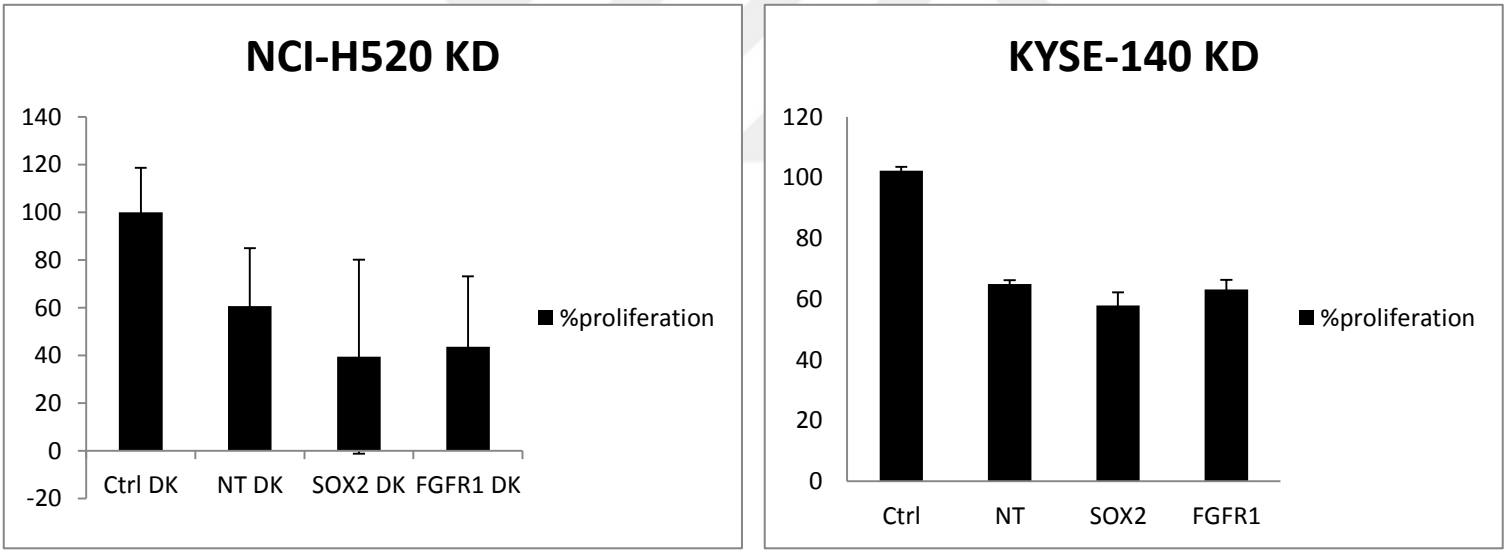


Fig .8. (A-B) Proliferation assay performed on NCI H520 cells ($1x10^4$ c/ml) and KYSE 140 cells ($1x10^4$c/ml) transfected with NT, SOX2 or FGFR1 targeting siRNA with a concentration of 20nM. Cells were plated triplicate on 96 well plates (200µl/well). siRNA treatment was performed twice for a double knockdown, and the MTT assay was done 48hrs subsequent to the knockdown procedure,

KYSE 140 cells with SOX2 knockdown showed ~44% decrease in proliferation ($p<0.05$) and *FGFR1* knockdown showed ~40% decrease in proliferation ($p<0.05$) relative to UT KYSE cells. On the other hand, both KYSE and NCI H520 cells also showed significant decreases in proliferation when transfected with Non-targeting (NT) siRNA pool; 38% drop in KYSE 140 cells and 40% decrease in NCI H520 cells (Figure 8).NT siRNA pool is composed of 4 preparations of scrambled siRNA designed unable to target any region in the genome so the real cause of this overt toxicity was unknown, but I sought to identify which siRNA in the NT pool was causing this toxicity by ordering each NT siRNA from the pool separately and performing an NT toxicity test by transfecting KYSE 140 cells with NT siRNA and NT pool in a MTT proliferation assay.

I performed a single transfection of NT pool and 4 NT siRNA of the smartpool separately, accompanied by UT and *GAPDH* knockdowns. Cells were incubated for 48hrs after transfection and assayed for proliferation via MTT assay. NT smartpool did not display the overt toxicity it showed on previous assays; with only a ~11% reduction in proliferation compared to UT cells. This could be due to transfection performed once and incubation time being only 48 hours, in comparison to double knockdown and 72 hours incubation time of knockdown assays. None of the separate NT siRNA transfections had significant toxicity on KYSE cells, but I chose NT #4 siRNA as having unnoticeable effect on cell viability and proceeded to use NT #4 in further proliferation and knockdown assays (Figure 9).

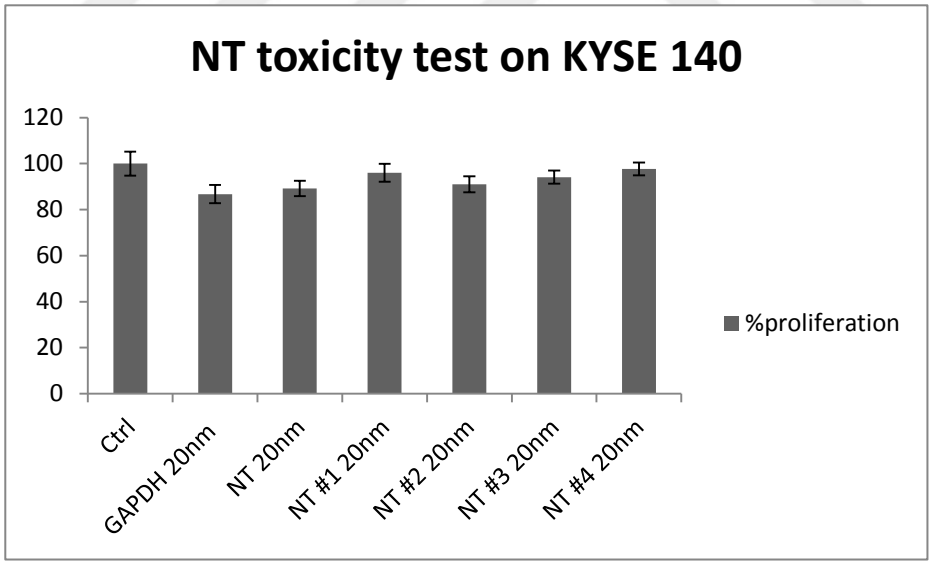


Fig .9. Proliferation assay done to assess the toxicity of siRNA nontargeting Smartpool elements on KYSE 140 cells ($1x10^4$ c/ml). Cells in each condition were transfected with siRNA two consecutive days with a siRNA concentration of 20nM. Condition designated NT is the smartpool composed of 4 different nontargeting siRNA fragments, whereas conditions designated NT #1-4 are transfected with the individual smartpool siRNA fragments.

Along with RNAi and Proliferation experiments, cells transfected with siRNA were observed at 48 and 72 hours after transfection via light microscopy for any changes in cell morphology arising from *SOX2* or *FGFR1* inhibition. I observed changes in KYSE 140 and NCI H520 cells with SOX2 and FGFR1 depletion consistent with autophagy. Vacuolated cells and granulated cells indicating the presence of autophagosomes were observed in *SOX2* and *FGFR1* knockdown KYSE 140 cells in greater quantities than UT, NT or *GAPDH* knockdown cells (Figure 10). On NCI H520 cells with depleted SOX2 or FGFR1, I observed decreases in the size of cell clusters and vacuolated cells in the process of cell death and subsequent detachment from culture surface (Figure 11).

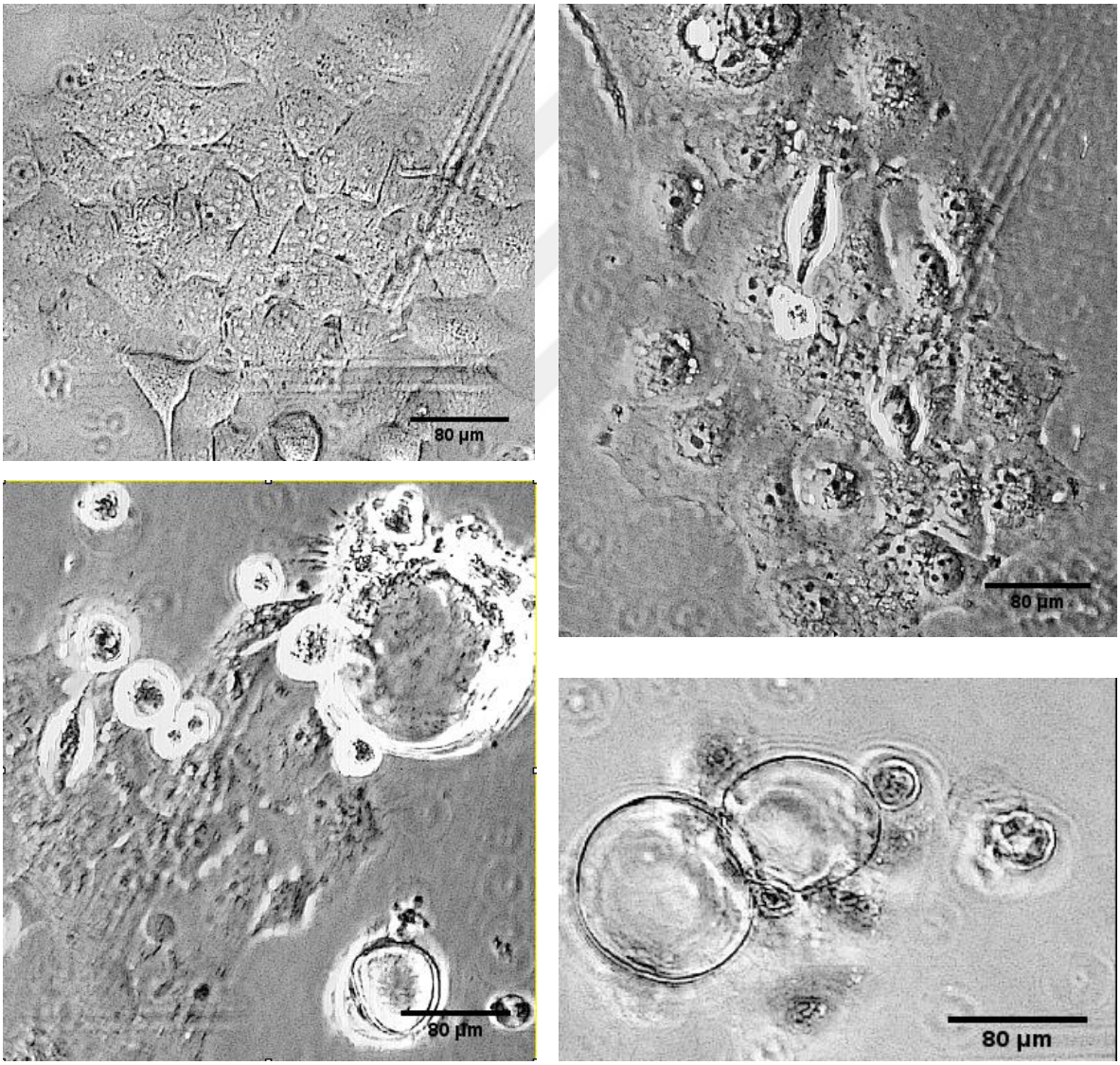


Fig .10. Microscopy Images taken from KYSE 140 cells with SOX2 and FGFR1 knockdown (40x Magnification on all images). **(A)** Untransfected KYSE 140 cells **(B)** KYSE 140 cells transfected with NT siRNA **(C)** KYSE 140 cells with FGFR1 knockdown. **(D)** KYSE 140 cells with SOX2 knockdown.

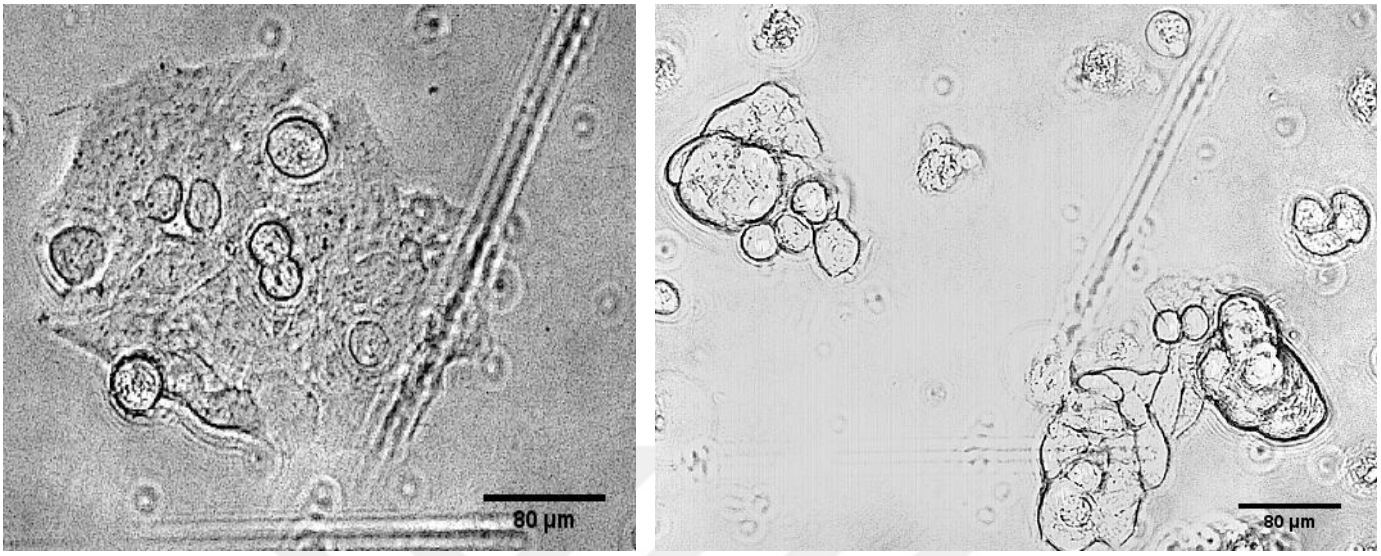


Fig.11. Microscopy Images taken from NCI H520 cells with SOX2 knock-down (40x Magnification on all images). **(A)** Untransfected H520 cells **(B)** H520 cells transfected with SOX2 targeting siRNA.

To correlate these observances with autophagy and presence of autophagosomes, cell lysates harvested from *SOX2* and *FGFR1* knockdowns were stained with LC3B antibody in western experiments. However; no evidence of LC3B flux was found on *SOX2* and *FGFR1* knockdown cells. This could be due to problems with antibody specificity or antibody degradation and time constraints prevented me from repeating western experiments with new LC3B antibody. Despite this failure, preliminary data has demonstrated that autophagy was induced in lentiviral *SOX2* knockdown by showing LC3B flux in transduced SQC cells (Figure 12).

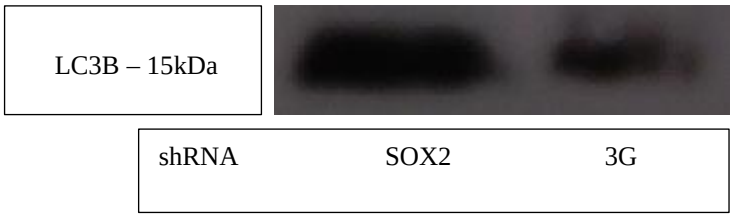


Fig .12. Preliminary data showing LC3B flux in SQC cells transduced by SOX2 targeting shRNA, using cells transduced by pLKO.3G empty vector as comparison.

Based on this preliminary data suggesting autophagy initiation on SOX2 depleted SQC cells, I decided to explore the effects of autophagy inhibitors on *SOX2* knockdown cells. It could be possible to identify a synergistic effect of autophagy inhibitors with SOX2/FGFR1 depletion, thus resulting in a greater reduction in cancer cell proliferation and increase cell death. By inhibiting a survival mechanism exploited by cancer cells, a demonstration of autophagy inhibitors suppressing a cancer survival pathway could be influential for further studies on this field. Towards that aim, I decided to treat *SOX2*/*FGFR1* knockdown cells with Chloroquine Bisphosphate (CQ) an anti-malarial drug with demonstrated autophagy inhibiting effect; and capable of being administered into human metabolism to a certain dose.

To determine the optimal dose of CQ to be given to cells, a dose response experiment was performed with CQ concentrations from 0µM to 100µM on KYSE 140 cells plated on 96 well plates in triplicate. Proliferation of cells was assayed via MTT with treatment times of 0, 48 and 72 hours. The aim was to determine a CQ dose sufficient to have an effect on cell viability but not causing overt toxicity which would mask any synergistic effect with SOX2 or FGFR1 depletion.

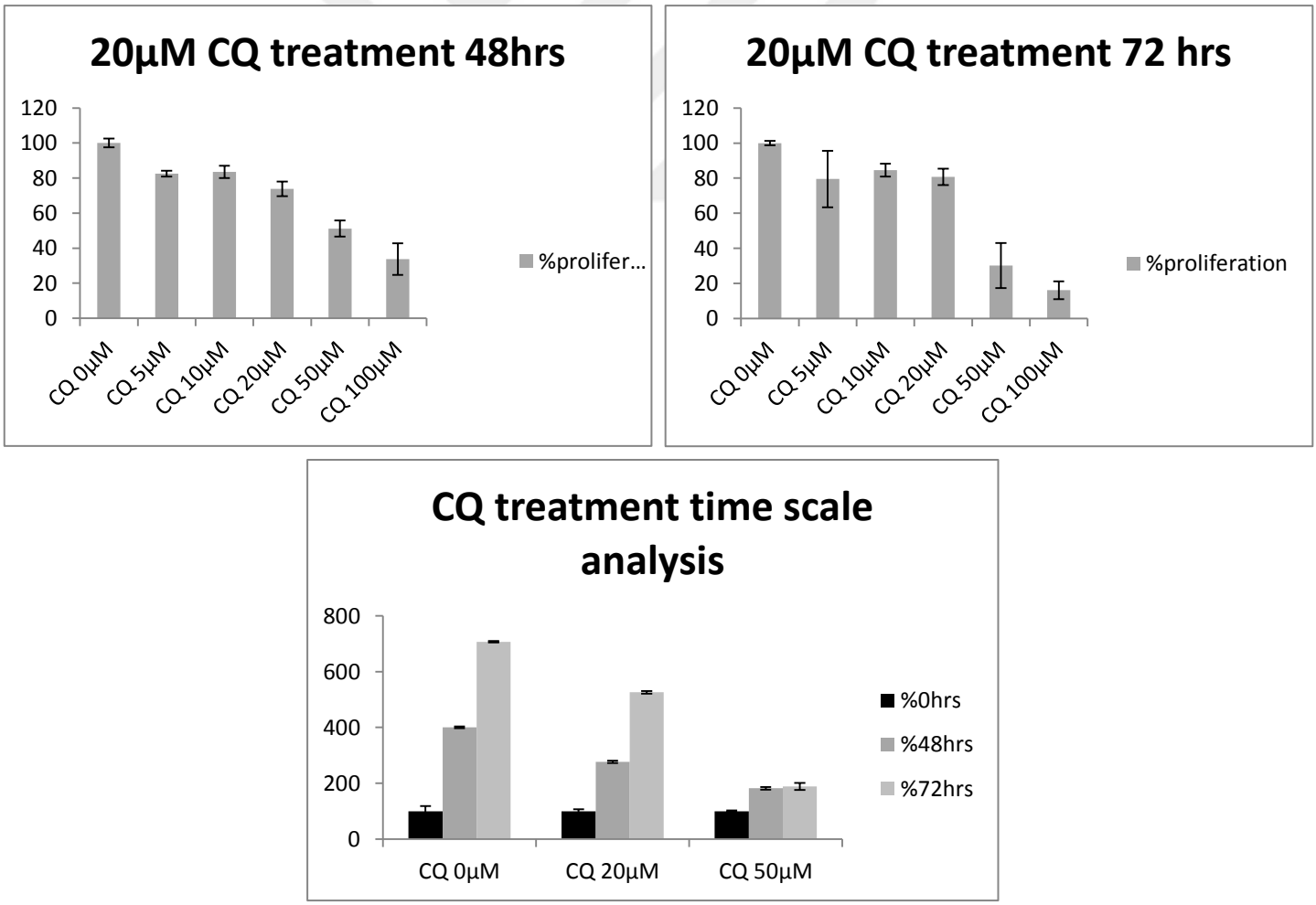


Fig .13. Dose response analysis of KYSE 140 cells 48 and 72 hours after treatment with CQ concentrations ranging from 0µM to 100µM **(A)** Dose response analysis of KYSE 140 cells 48 hours after treatment with CQ **(B)** Dose response analysis of KYSE 140 cells 72hrs after treatment with CQ. **(C)** Time scale analysis of KYSE 140 cells treated with CQ concentrations 0, 20 and 50µM at 0, 48 and 72hrs after treatment.

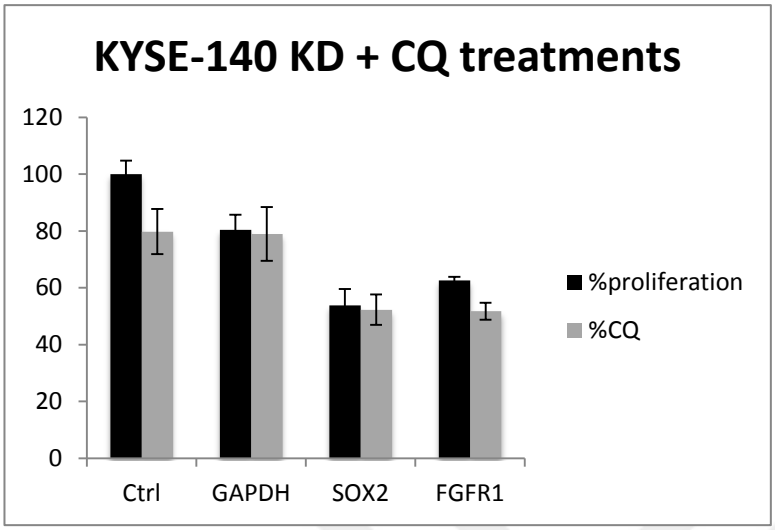
The results of the proliferation experiment showed that doses exceeding 20μM CQ would result in significant toxicity and decrease in proliferation. At 48 hours after treatment, 50μM CQ would decrease proliferation by ~50% compared to untreated cells ($p<0.005$) while 20μM CQ would only result in ~23% reduction in proliferation ($p<0.05$) (Figure 12). Further incubation times would result in further proliferation drops (~70% at 72 hours) at doses higher than 20μM (Figure 13) Also studies using CQ as an *in vitro* drug commonly use doses ranging from 10μM to 20μM. Thus; I decided to administer 20μM CQ concentration with an incubation time of 24 hours in further assays combined with siRNA knockdown.

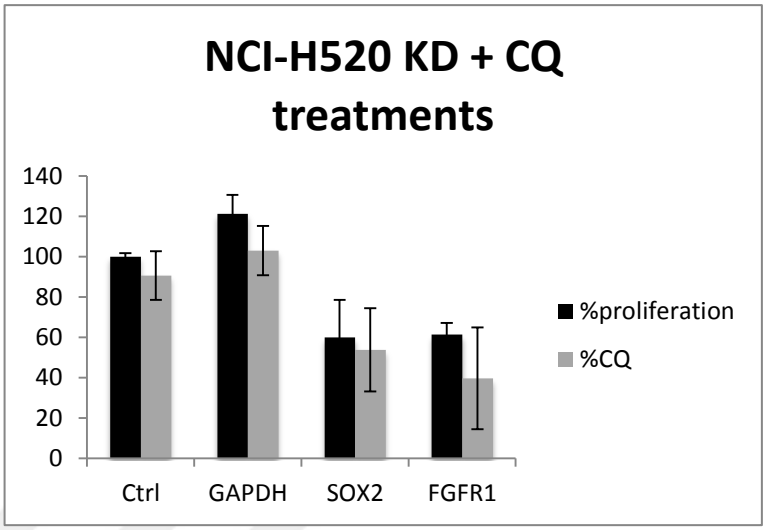
At the last stage, the aim was to demonstrate evidence of a possible synergistic effect on double knockdown of *SOX2* or *FGFR1* followed by a CQ treatment on cell death. I also set up experiments with two additional cell lines: LUDLU1 is an SQC cell line with normal copy number in *SOX2*, and BEAS-2B an immortalized human normal lung epithelial cell line with normal copy numbers in SOX2 and FGFR1. These cell lines would be used to compare effects of SOX2/FGFR1 depletion and CQ treatment with the effects on cells with high level *SOX2*/*FGFR1* amplification.

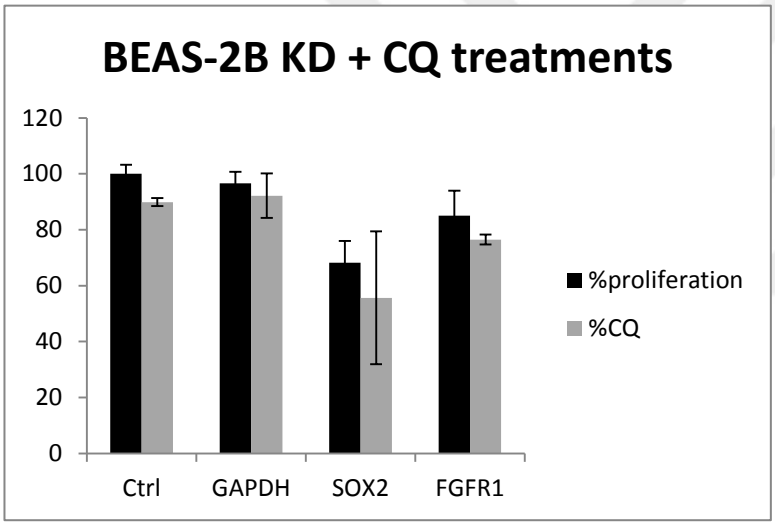
I set up proliferation assays with the 4 different cell lines on 96 well plates with each condition in triplicate; performing double knockdowns of SOX2 and FGFR1 with 20nM siRNA, and treating them with 20μM CQ with 24 hour intervals between each procedure. 24 hours after CQ treatment (72 hours after first siRNA transfection) I performed MTT assays to assess the effects of this combined treatment on my cell lines. However; multiple manipulations upon cells growing on 96 well plates introduced bias to my assays on many occasions, including cell desiccation, which prevented me from acquiring meaningful data on many assays. I had two successful sets of experiments with UT and untreated cells were unaffected by siRNA/CQ implementations.

I acquired a ~11% reduction in proliferation ($p<0.05$) in KYSE 140 cells with *FGFR1* KD and CQ treatment compared to cells with only FGFR1 depletion. *FGFR1* KD + CQ treated cells also showed ~28% decrease in proliferation ($p<0.05$) when compared to UT KYSE cells with only CQ treatment. In contrast I did not observe a significant reduction in proliferation in *SOX2* KD + CQ treated KYSE cells in comparison to *SOX2* knockdown cells. Both conditions demonstrated a decrease in proliferation ~50% ($p<0.05$) when compared to UT cells.

On NCI H520 cells with *FGFR1* KD + CQ treatment I observed a ~22% further decrease in proliferation than H520 cells with only *FGFR1* inhibition. However the variation in absorbance values were too high for the Student's p-test to return any significant change ($p=0.109$).Similar to KYSE 140 cells, H520 cells with *SOX2* KD + CQ treatment did not undergo a further reduction in proliferation in comparison to *SOX2* KD cells.







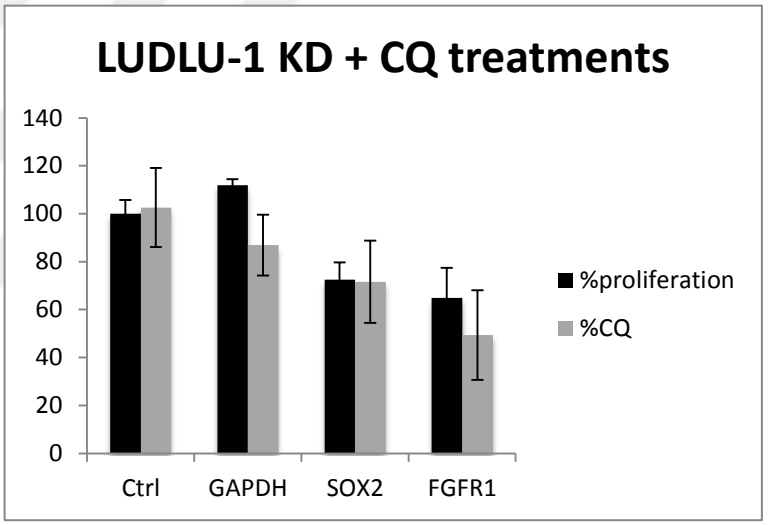


Fig.14. (A-D) Proliferation assay performed on NCI H520 cells ($1x10^4$ c/ml) KYSE 140 cells ($1x10^4$c/ml) LUDLU1 ($1x10^4$c/ml) and BEAS 2B ($1x10^4$c/ml) cells transfected with NT, SOX2 or FGFR1 targeting siRNA with a concentration of 20nM, with or without 20µM CQ treatment 24hrs after siRNA transfection. Cells were plated on 96 well plates (200µl/well) each condition in triplicate. siRNA treatment was performed twice for a double knockdown, and the MTT assay was done 24hrs subsequent to the CQ treatment.

LUDLU1 cells with *FGFR1* KD + CQ treatment resulted in a ~15% further decrease ($p<0.05$) in cell proliferation than LUDLU cells with only FGFR1 knockdown. There was no difference in proliferation values between SOX2 KD+ CQ treated and only SOX2 knockdown cells. *SOX2* KD and *SOX2* KD+ CQ cells had ~29% reduction in proliferation ($p<0.05$) when compared to UT cells. *FGFR1* KD cells showed ~36% decrease in proliferation ($p<0.05$) and *FGFR1*+ CQ treated cells had ~51% reduction in proliferation ($p<0.05$) in comparison to UT cells.

On BEAS-2B cells; ~15% reduction ($p=0.11$) was observed in *FGFR1* KD cells, and 32% proliferation reduction ($p<0.05$) was calculated in *SOX2* KD cells. In both knockdowns, CQ treatment did not result in a further drop in proliferation.

In the end, data acquired from these experiments suggest that an additive could be present between FGFR1 depletion and autophagy inhibition by CQ in SQC cells. However I did not observe any evidence for same degree of synergy between CQ and SOX2 knockdown. Given the fact that the scope and number of sets of experiments is limited, resulting in a small database, further experimentation with the same techniques will be required to uncover any synergy effect of CQ with SQC oncogenes and clarify the extent of synergy that I observed between *FGFR1* and CQ.

3.2 Interrogation of LKB1 and P16 on SQC via the MCC technique

10 sets of Copy number primer sets and 5 sets of Mutation primer sets (each corresponding to the 5 mutation hotspots in *LKB1*) were designed for *LKB1*. 6 sets of copy number primer sets were designed for p16.

Primer sets designed for use in digital PCR experiments were first validated by tests on human normal lung epithelial DNA. In normal human DNA, the prediction would be that all autosomal markers (which include LKB1 and p 16, alongside reference markers) will give the same copy number; allowing us to compare the relative copy number from the gel electrophoresis results of each primer set to a series of previously tested reference markers (McCaughan et al. 2010)

Primer sets with calculated copy number similar to my reference markers would be retained for further digital PCR tests, while primer sets with significantly lower band counts or no target amplification would be eliminated. Lower number of positive bands in comparison to reference markers would imply a relatively low copy number of target gene on normal DNA, which would indicate the primer set is incapable of reflecting the copy number of the target gene.

For validation of mutation primer sets, the relative copy number is less important than the presence of amplification (any number of positive bands); since the mutation marker is designed to amplify a specific site along the *LKB1* gene, not give information about the copy number of the gene itself.

Some of the mutation primers designed could be used both as a copy number set and a mutation set, which is ideal for use in both sequencing and copy number calculations, but given the constraints for designing mutation markers (Methods, Primer Design) these dual use primers comprised only a small portion of total primer sets designed.

After the validation of primer sets; from which I present examples of working primer sets and eliminated primer sets compared to a reference marker via gel electrophoresis images (Figure 15); 2 out of 6 (CN3 and CN6) copy number markers for P16, 2 out of 10 (CN1 and CN9) copy number markers for LKB1, and 3 out of 5 (M1, M4 and M6) mutation markers for LKB1 were validated for further use in MCC protocols.

To increase the number of primer sets for use in LKB1, I designed 5 more copy number markers and re-designed 3 mutation markers for the 2 hotspots that failed to work to keep the scope of my experiments to encompass all mutation hotspots along LKB1. From the new copy number primers, 2 out of 5 (CN11 and CN14)were validated for further use. 2 out of 3 (M2 and M3) of the mutation markers were successful in amplifying their corresponding mutation sites. The initial primer sets and those that were validated as working can be found in Tables 3, 4 and 5.

After the validation tests; I had 4 copy number primer sets on LKB1, and 2 copy number sets set on P16 and 5 mutation primer sets for mutation points on LKB1. On the other hand even one copy number set would be sufficient to reflect the state of the gene in a lesion. Thus, I had more copy number sets designed than required for a safe calculation of copy number.

Table 3. Initial Copy number markers for P16

Marker Name	Locus	Fex	Fin	Rvs	Amplimer Start $	Amplimer End$	Amplimer Length
P16CN1	P16	tcgtgctgatgctactgagg	aagaccaggtaggaaaggcc	Acaaaacaagtgccgaatgc	26428	26358	109
P16CN2	P16	aaggtgcctcgtgttaggc	gcccttcttgggatctcttc	Tggaggggactagagtgtgc	22466	22397	107
P16CN3	P16	tattcctccattgcctttgc	ttgcacccaacatcctattct	Atggagctcccaggtacagc	13763	13697	105
P16CN4	P16	caacgcaccgaatagttacg	atccaggtgggtagagggtc	Aattcccctgcaaacttcg	6944	6874	108
P16CN5	P16	gggaaataatcccgaaatgg	aaagcataccaccacccaaa	Aggccttgaactagcagagg	2954	2885	108
P16CN6	P16	ctgccttttcactgtgttgg	ttctggagtgagcactcacg	Tcatgaagtcgacagcttcc	214	140	113
P16CN9	P16	tgccctgcttttactgttcc	cttttcctgtctcccagctg	Ccctttttcttccacatcacc	27900	27796	105
P16CN10	P16	tctagggggacctcatatcg	actggcacatctggagatcc	Gcctggtgagcaaaatatcc	-200	-310	111

$: Base Pair numbers correspond to the P16 nad LKB1 files retrieved from ENSEMBL. Base 1 of LKB1 gene corresponds to the 1205798 numbered base along Chr 19. Base 1 of P16 gene corresponds to the 21967751 numbered base along Chr 9.

Table 4. Initial Copy number and mutation markers for LKB1

Marker Name	Locus	Fex	Fin	Rvs	Mutation Type	Amplimer Start $	Amplimer End $	Amplimer Length
LKB1M1.1	LKB1	Atccaccgcatcgactcc	aagtacctgatgggggacct	tcaccttgccgtaagagc	c.109C>T	1191	1305	115
LKB1M2.1	LKB1	Ccgcaggtacttctgtcagc	gtcagctgattgacggcc	ccaggtcggagattttgagg	c.508C>T	14569	14696	128
LKB1M3.1	LKB1	Ggaacctgctgctcaccac	accaccggtggcacc	gtgcagccctcagggagt	c.580G>T,c.595G>T	14649	14769	121
LKB1M4.1	LKB1	Ttcgaaggggacaacatctac	gtttgagaacatcgggaagg	gatgaggctcccacctttc	c.842C>T	15443	15556	114
LKB1M4.2	LKB1	Cttcgaaggggacaacatctac	gtttgagaacatcgggaagg	gatgaggctcccacctttc	c.842C>T	15442	15556	115
LKB1M4.3	LKB1	Ttcgaaggggacaacatctac	gtttgagaacatcgggaagg	atgaggctcccacctttcag	c.842C>T	15443	15555	113
LKB1M5.1	LKB1	Catgactgtggtgccgtact	cgtacttggaggacctgca	caccgtgaagtcctgagtgt	c.842C>T	17268	17370	103
LKB1CN1	LKB1	Tagggaagggaggaggtacg	acttccacagggagatggg	gaggacaggggtgtatcagc	c.1062C>G	12503	12608	106
LKB1CN2	LKB1	Aaaggccacacaatgtaccc	agtgactcaagggtggcctt	ggaggatggagaagctgagg		21613	21726	114
LKB1CN3	LKB1	Ggtcactgtctgccatcagg	ccgcttacagctgtgattca	acaccagctcccagaatagg		4266	4380	115
LKB1CN4	LKB1	Gaactggggtctgagtgagg	cacagctggtcccaaacac	gaggaggggaaaaagagagg		16413	16526	114
LKB1CN5	LKB1	Ctccaccgaggtcatctacc	aagtacctgatgggggacct	agcacctccttcaccttgc		1205	1315	111
LKB1CN6	LKB1	Ttcatcctgctctcctagcc	agccctggcagagctcag	caacctcagaaggccaagg		9704	9812	109
LKB1CN7	LKB1	Tggtgaagacagaggtgtcc	gtagagctggggctcctagg	gtcagtcaaggtggttgacg		15270	15383	114
LKB1CN9	LKB1	Ggcaactcttgtttttcacg	atgagttccgatagggcaga	gggaaagagaaacgctcagg		-663	-551	113
LKB1CN10	LKB1	Aggagcccgaagctgagag	gttcacgccccttccttc	actgctctttccggtgcttg		23065	23174	110
LKB1CN11	LKB1	Ggcctctgctgtctgaaaac	caccgcatctcctctgaatag	atgaaggtcacaggccattc		9135	9238	104
LKB1CN12	LKB1	Cagggtcttctgcctttcag	caggagggccagttaggg	cagtagggactctcgcaacc		18652	18767	114
LKB1CN13	LKB1	Gtcaggcttggagtcaggtc	cctagaggacatggctgagc	agggtacctgccacacactc		16990	17106	117
LKB1CN14	LKB1	Agggtgtgtacctgggagtg	cctttggagaagctgcagac	gaggagcaaggtcctcacac		11293	11399	107
LKB1CN15	LKB1	Tgcatccggaagaccttact	actgacttcttcaggcacgg	gctgccctaggtaggagctt		2037	2147	111

Table 5. Validated copy number and mutation markers for LKB1 and P16

Marker Name	Locus	Fex	Fin	Rvs	Mutation type	Amplimer Start $	Amplimer End $	Amplimer Length
LKB1M1.1	LKB1	tcaccttgccgtaagagc	aggtcccccatcaggtactt	Atccaccgcatcgactcc	c.109C>T substitution	1305	1191	114
LKB1M2.1	LKB1	ccgcaggtacttctgtcagc	gtcagctgattgacggcc	Ccaggtcggagattttgagg	c.508C>T substitution	14569	14696	128
LKB1M3.1	LKB1	gggtgcagtgcctgtgg	gtgcctgtggcggtgc	Tggcaccctcaaaatctcc	c.580G>T, C.595G>T substitutions	14793	14671	122
LKB1M4.1	LKB1	ttcgaaggggacaacatctac	gtttgagaacatcgggaagg	Gatgaggctcccacctttc	c.842C>T substitution	15443	15556	114
LKB1M6.1	LKB1	catgactgtggtgccgtact	cgtacttggaggacctgca	Caccgtgaagtcctgagtgt	c.1062C>G substitution	17268	17370	103
LKB1CN1	LKB1	tagggaagggaggaggtacg	acttccacagggagatggg	Gaggacaggggtgtatcagc	CN	12503	12608	106
LKB1CN9	LKB1	ggcaactcttgtttttcacg	atgagttccgatagggcaga	gggaaagagaaacgctcagg	CN	-663	-551	113
LKB1CN11	LKB1	ggcctctgctgtctgaaaac	caccgcatctcctctgaatag	Atgaaggtcacaggccattc	CN	9135	9238	104
LKB1CN14	LKB1	agggtgtgtacctgggagtg	cctttggagaagctgcagac	Gaggagcaaggtcctcacac	CN	11293	11399	107
P16CN3	P16	tattcctccattgcctttgc	ttgcacccaacatcctattct	Atggagctcccaggtacagc	CN	13763	13697	105
P16CN6	P16	ctgccttttcactgtgttgg	ttctggagtgagcactcacg	Tcatgaagtcgacagcttcc	CN	214	140	113

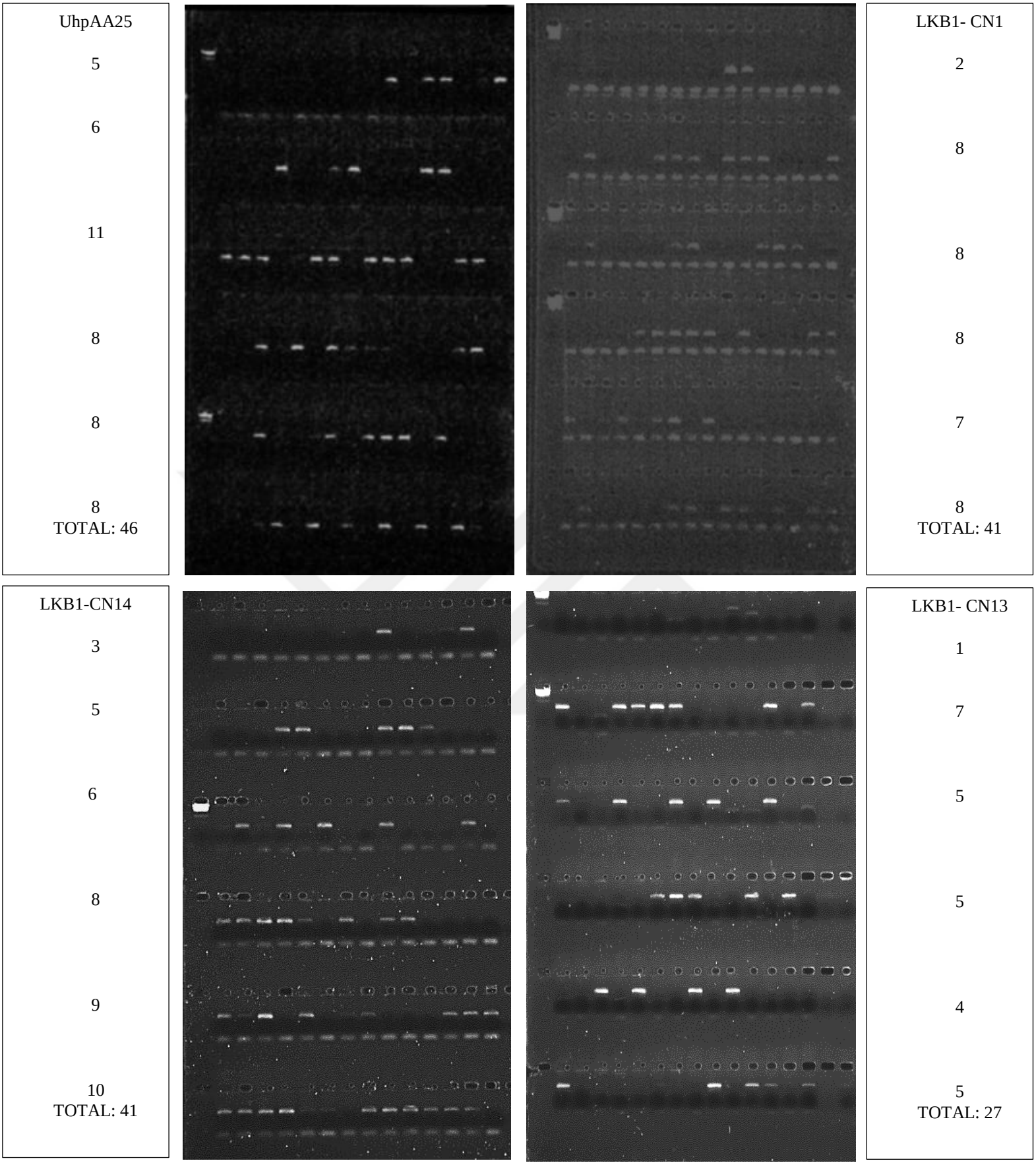


Fig. 15. Sample gel electrophoresis photos for validating LKB1 copy number primer sets **(A)** Reference image representing a gene with normal copy number. Each primer set is required to give a relative copy number similar to reference marker to pass the validation test .**(B-C)** Copy number sets 1 and 14 for LKB1 reflect the copy number of LKB1 in human normal lung epithelium genome. **(D)** Copy number set CN13 shows band count of a significantly lower value compared to CN1, CN14 or reference gene, and is eliminated from the primer list.

Reproduction of µMCC Technique on DNA from Low and High G Lesions

The first copy number test for LKB1 and P16 was done on DNA extracted from a paraffin embedded biopsy sample of a lesion histopathologically defined as "low grade lesion". Relative copy number of the target genes were calculated with the band number of positive aliquots returned from the pool of copy number primer sets; using the Poisson distribution formula described in Methods (Equation 1). Relative copy number of LKB1 and P16 was compared to previously tested reference markers (Figure 15).

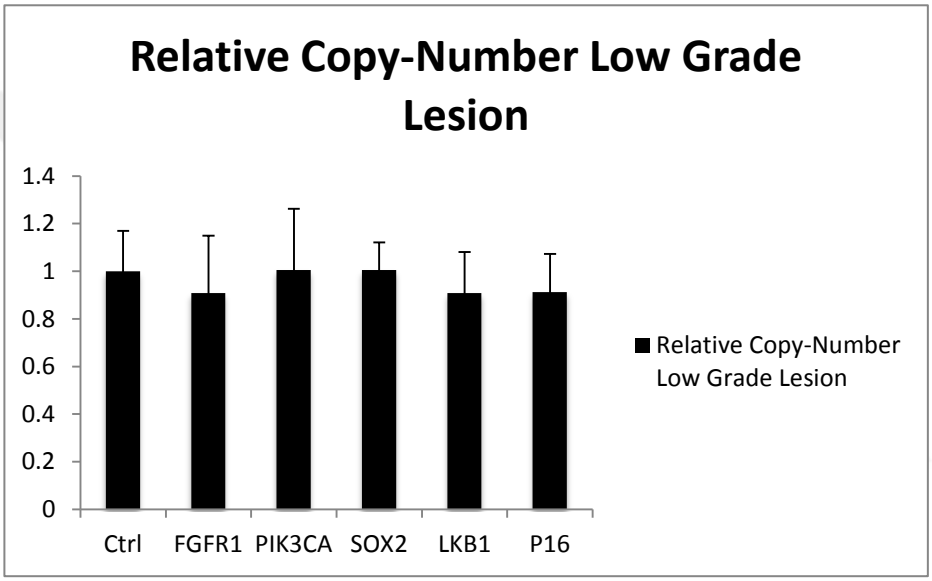
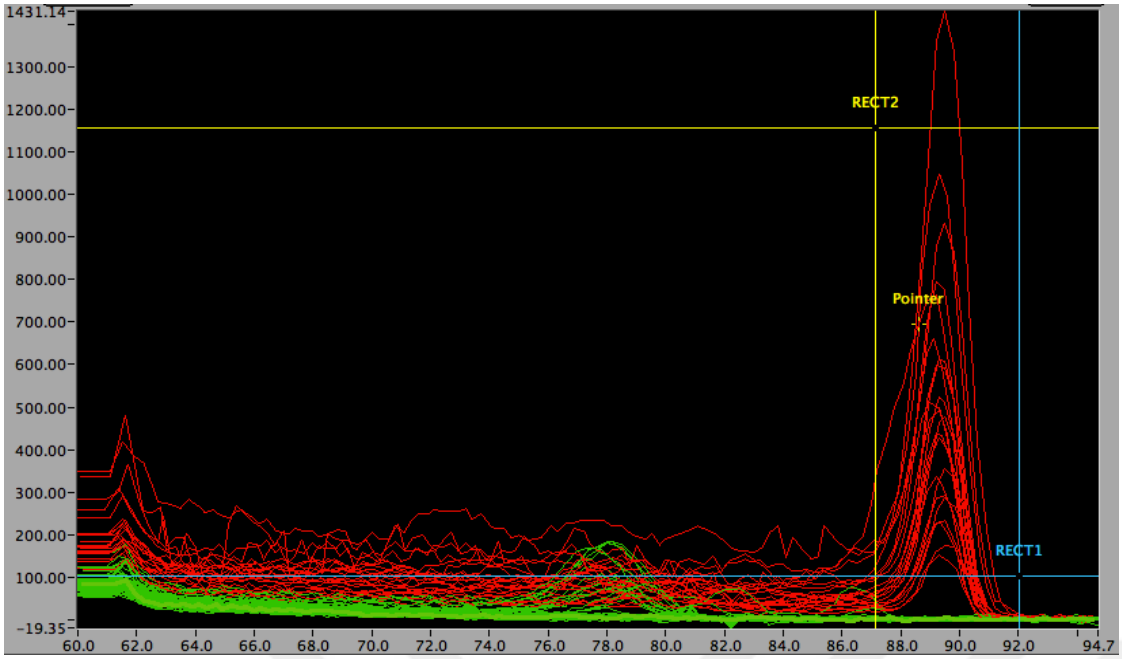
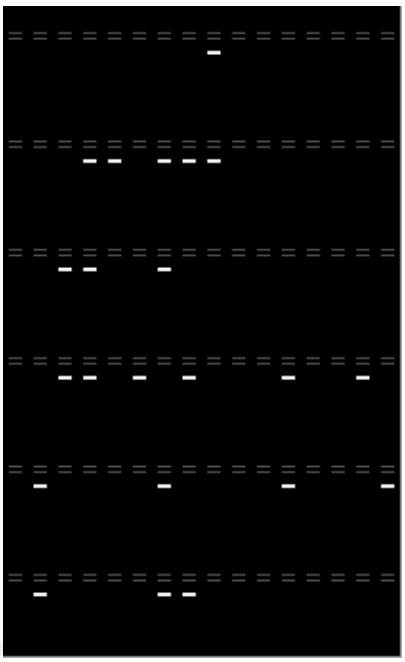
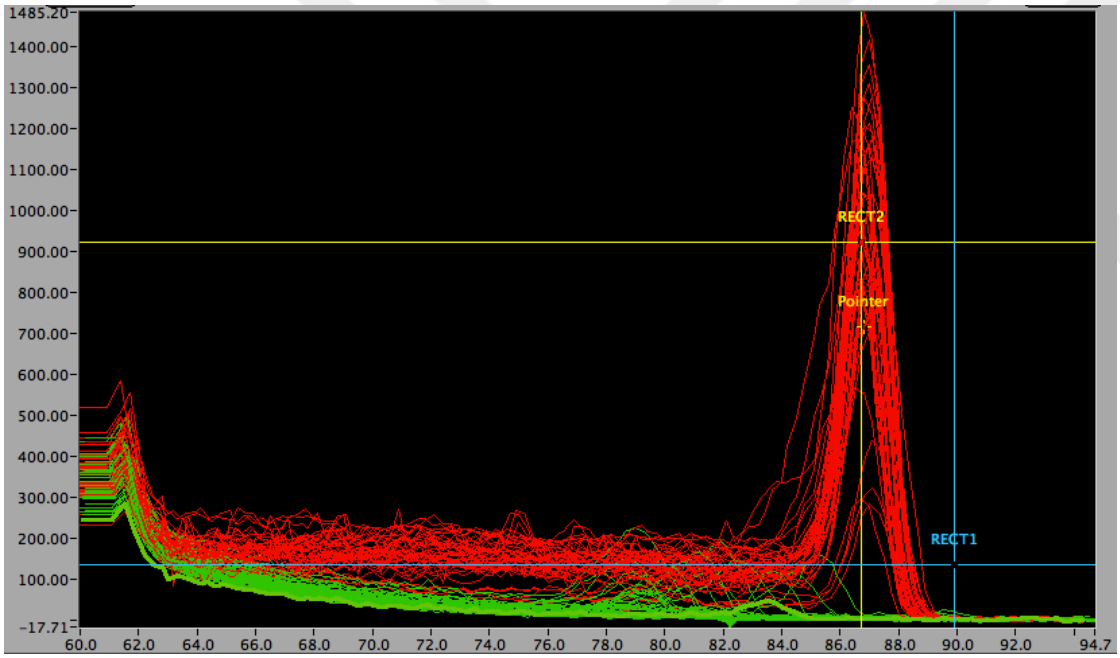


Fig. 16. Relative Copy Numbers of LKB1 and P16 genes in comparison to reference markers commonly amplified in SQC lesions.

On this lesion, LKB1 and P16 relative copy numbers were calculated to be 1 (Figure 15).This indicates that both tumour suppressor genes were not yet lost in this stage of development for SQC. It was also seen that the reference markers also tested on this early lesion; SOX2, FGFR1 and PIK3CA were also at normal copy number. These are oncogenes that are documented to be commonly amplified in copy number in late stages of development of SQC.







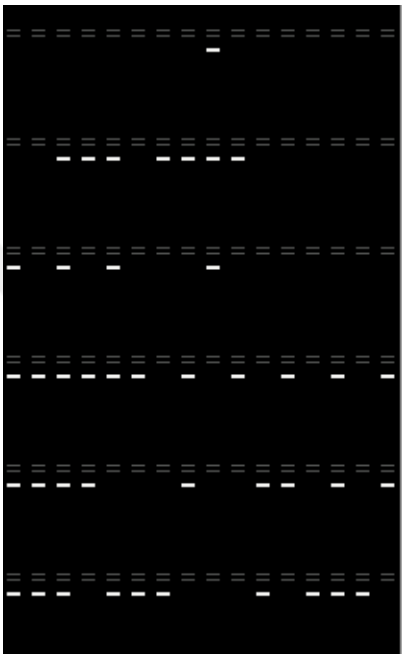


Fig. 17. Melting Curve analysis images and digitally constructed representative gel images of LKB1 mutation primers M1 and M2. **(A, C)** Melting curves of each positive aliquot is superimposed to calculate the melting temperature of each amplified species **(B,D)** Positive wells are clearly identified by the representative gel image constructed by positions of positive aliquots. Images retrieved from KingCurve software.

I then tested my LKB1 mutation markers on the same DNA extracted from low grade SQC lesion. The results were analysed by melting curve analysis, of which two of the generated melting curves for positive aliquots and the digital representative gel electrophoresis images are presented here (Figure 17). Mutation marker 2 set giving a relative copy number of 1 (42 positive aliquots) indicates that this primer set can be used also as a copy number set in a dual role.

M13 sequencing tags were added onto mutation markers and the digital PCR tests were re-ran the on Low grade dysplasia DNA, sending the results of this test to sequencing to analyse the presence of possible mutations on each of the 5 mutation hotspots on LKB1 in low grade lesion genome.

I was able to analyse 4 out of 5 of the mutation hotspots on LKB1. Sequencing results for mutations 1, 2, 4 and 5 returned precise sequences of the amplimers generated via digital PCR tests (Figure 18). However the sequencing reaction for mutation 3 failed to return any results. No mutations were present in any of the mutation hotspots recorded along LKB1. This is not surprising as the lesion analysed is histopathologically low grade dysplasia, and is thought to have a relatively stable genome compared to lesions in later stages of SQC development. Also the mutations documented on COSMIC database are present only on ~20% of total analysed SQC lesions. It would be an extremely rare event to document a mutation from the sequencing analysis of only one lesion.

On the other hand, this test have provided ample evidence that the Mutation markers I designed are capable of being used in sequencing and return sequencing results of high confidence, along with the capability of identifying low frequency mutation events, as they are designed for digital PCR applications, in contrast to primers designed for bulk sequencing applications.

On the high grade lesion I tested the copy number amplification of SQC oncogenes SOX2, FGFR1 and PIK3CA to include an example of gross amplification of an oncogene in late phase SQC lesions. Copy numbers of LKB1 and P16 were also tested on the high grade lesion but time constraints prevented me from analysing the data in a presentable way.

Melting curve images of SOX2 oncogene marker tested on high grade dysplasia clearly show the extent of copy number amplification encountered on the oncogene, when compared to the control marker representing a locus at normal copy number (Figure 20). Oncogene SOX2 shows an increase of copy number ~400% in comparison to the reference marker ($p<0.05$). Oncogene PIK3CA also demonstrates an increase in copy number reaching ~400% ($p<0.05$). In contrast the copy number calculations on FGFR1 showed that this oncogene is not amplified in copy number on this lesion, showing a relative copy number value of 1, similar to the reference marker (Figure 21).

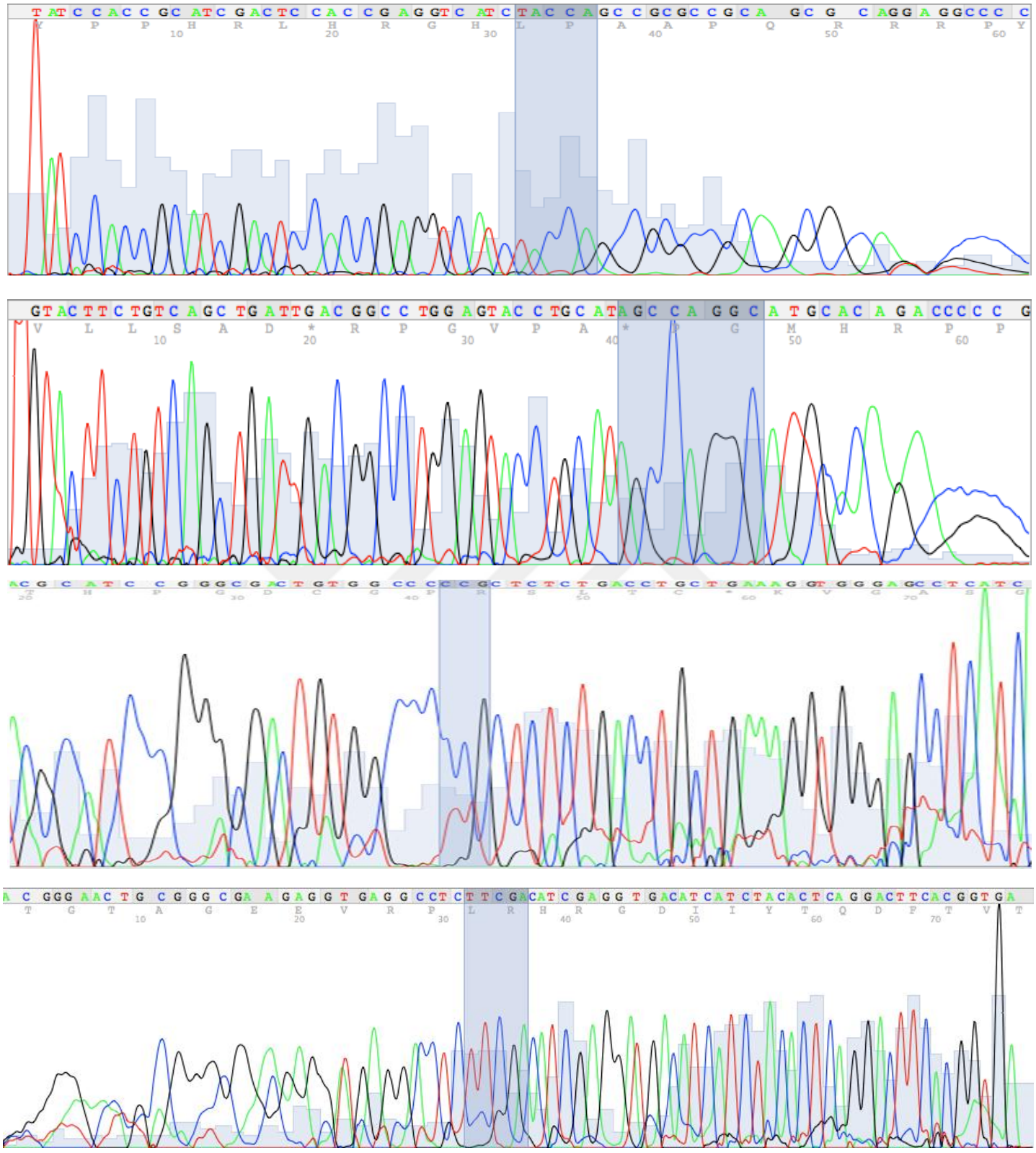


Fig. 18. Sequencing results for the 4 out of 5 mutation hotspots on LKB1 gene, sequenced by addition of M13 sequences to the ends of designed Mutation Markers for LKB1. Highlighted sequences are mutation sites and 1-5 adjacent base pairs flanking each mutation hotspot. Sequencing results for the Mut 3 sample failed to return any results.**(A)** Mut 1 c.109C>T substitution **(B)** Mut 2 c.508C>T substitution **(C)** Mut 4 c.842C>T substitution **(D)** Mut 5 c.1062C>G substitution

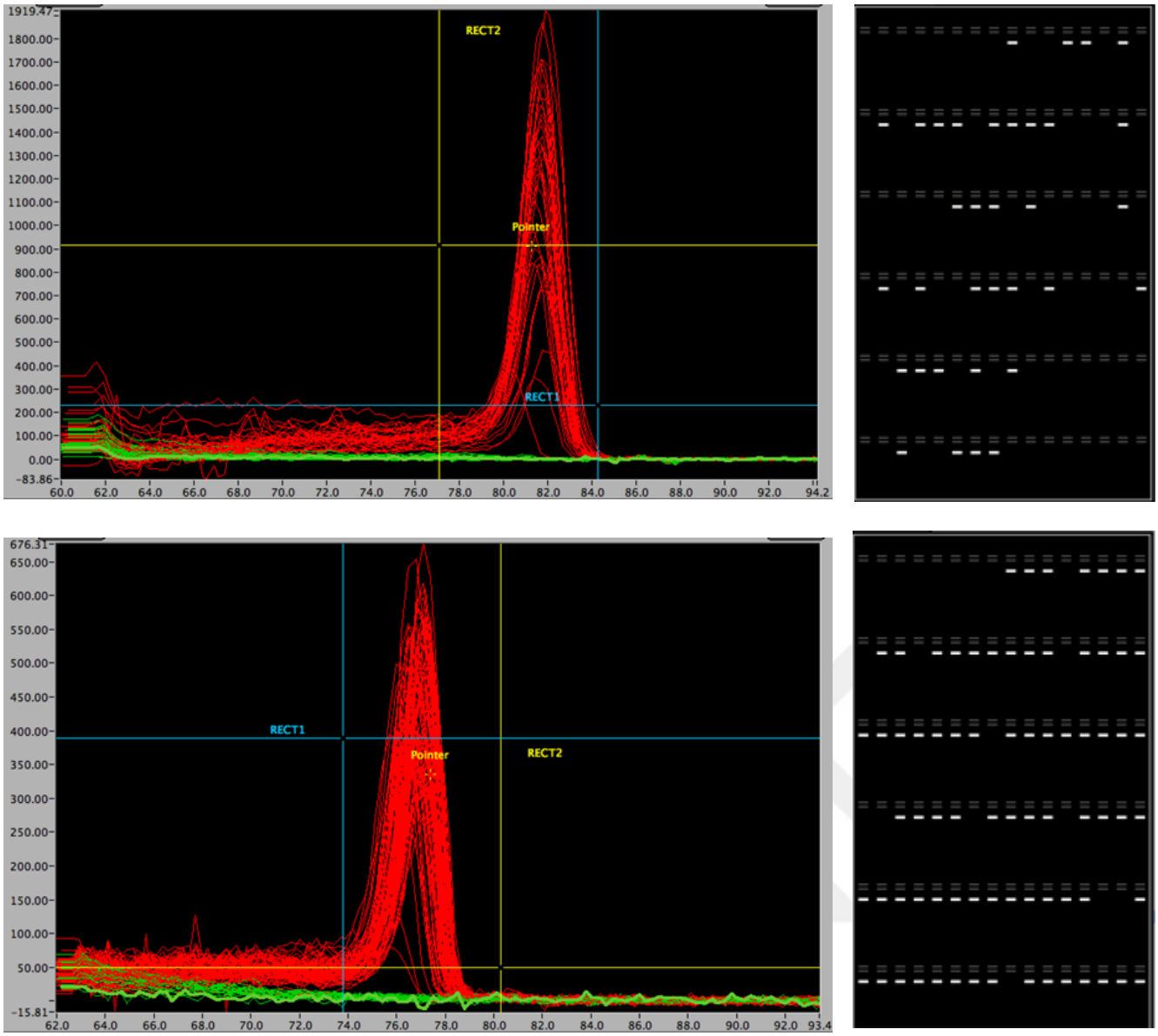
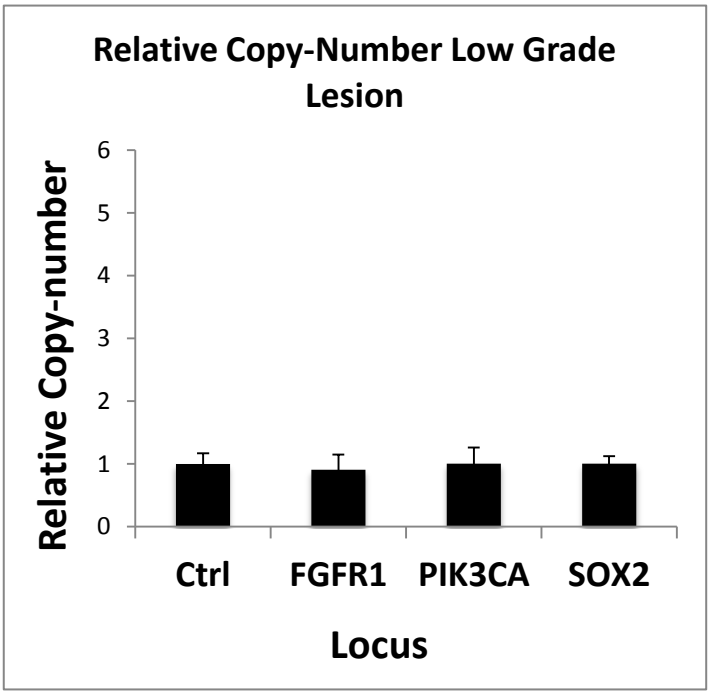


Fig. 19. Melting Curve analysis images and digitally constructed representative gel images of Control reference marker and SOX2 copy number marker tested on high grade lesion DNA. **(A, C)** Melting curves of each positive aliquot is superimposed to calculate the melting temperature of each amplified species **(B,D)** Positive wells are clearly identified by the representative gel image constructed by positions of positive aliquots. (Images retrieved from KingCurve software)



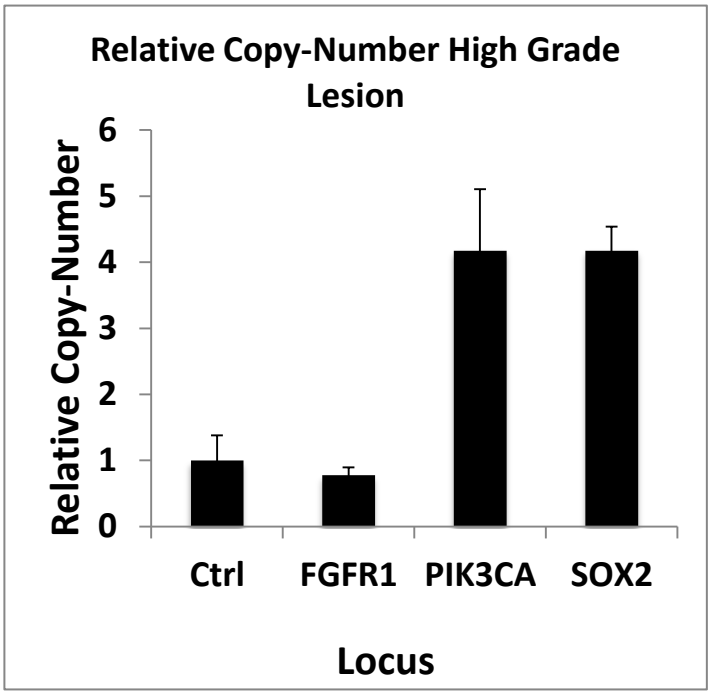


Fig. 20. (A) Relative copy number calculations of SQC oncogenes FGFR1, PIK3CA and SOX2 on DNA extracted from a low grade lesion **(B)** Relative copy number calculation of SQC oncogenes on High grade lesion DNA

It can be concluded from the comparison of tests on low grade dysplasia and high grade lesion, that SOX2 and PIK3CA oncogenes become amplified in copy number during the transition between low grade and high grade dysplasia, while FGFR1 copy number remains stable until after the formation of high grade lesion.

4 Discussion

4.1 RNAi Experiments on SQC cell lines

I have demonstrated successful knockdown of SOX2 and FGFR1 in a series of squamous carcinoma cell lines. This was confirmed by western blotting and resulted in a reduction in proliferation in treated cells compared to controls as measured using the MTT proliferation assay.

Further, consistent with preliminary data using a lentiviral shRNA system the knockdown of these genes appeared to induce the morphological appearances of autophagy. I was unable to confirm that autophagy was initiated on SQC cell lines on the wake of SOX2/FGFR1 depletion (for the lack of first hand data demonstrating LC3B flux upon siRNA transfection), and the morphology changes are insufficient on their own to confirm autophagy unless resolved by electron microscopy. To confirm the further decrease in proliferation stemming from a synergy between *FGFR1* knockdown and CQ, evidence of autophagy in FGFR1 depletion must be presented.

Subsequent attempts to establish whether SOX2 knockdown, in addition to chloroquine, an autophagy inhibitor, did not show a consistent or significant additive effect. This was in contrast to preliminary experiments using a shRNA approach.

On the other hand, further reductions in proliferation ranging from ~11% to ~22% were calculated in FGFR1 knockdown combined with CQ treatment compared to only FGFR1 knockdown cells. This may be a result of CQ inhibition eliminating initiation of autophagy caused by FGFR1 depletion, thus increasing cell death among cancer cells surviving the stress caused by *FGFR1* knockdown via autophagy.

It is often the case that siRNA and shRNA experiments do not exactly concur and the reasons for this are unclear but may be related to the duration of knockdown with the shRNA approach. We have generated fluorescently labelled shRNA lentiviral particles targeting SOX2 (pLKO.3G backbone) to compare directly with the siRNA used in these experiments.

It remains unclear, despite the convincing light microscopy pictures, whether or not autophagy was induced in this system. To clarify the presence or absence of autophagy in this system in the future a number of approaches could be taken including electron microscopy, using a fluorescently labelled LC3 to track autophagosome formation and resolution, and genetic inhibition of key autophagy pathway mediators.

The reason this may be important is that autophagy inhibitors, particularly CQ are now in clinical trials in patients with solid organ cancers. However the experimental data remains confusing with the inhibition of autophagy in certain experimental situations potentiating carcinogenesis.

Knockdown of SOX2 causes depletion of FGFR1

Knockdown experiments comparing FGFR1 protein levels in SOX2 knockdown and FGFR1 knockdown cells demonstrated a striking reduction of *FGFR1* expression in SOX2 depleted KYSE and H520 cells. Repeated experiments in my SQC cells with SOX2 amplification consistently gave the same results: SOX2 depletion was causing an observable reducing effect on *FGFR1* expression, or par with or even exceeding the effect of FGFR1 knockdown via siRNA (Figure 21).

As a transcription factor, SOX2 may be capable of binding to *FGFR1* locus in the genome and altering the expression of *FGFR1*, thus possibly contributing to SQC phenotype in a previously undocumented way. If high copy number of *SOX2* in SQC cell lines has an effect of increasing *SOX2* expression, a higher concentration of SOX2 may result in SOX2 binding to sites in cancer genome unused in normal genome.

Another possibility is that SOX2 might be acting in concert with other transcription factors directly responsible for *FGFR1*gene expression, and thus able to alter its transcription in an indirect way. It is known that SOX2 forms a trimeric complex with OCT 3/4 and capable of initiating expression of many genes critical in embryonic differentiation and pluripotency of esophageal stem cells.

Validity of this observation can be discerned by performing ChIP assays on SOX2 binding regions across human genome and investigating if SOX2 is capable of binding to *FGFR1* locus or other transcription factors that directly alter FGFR1 expression such as CREB, ΔCREB, SRF or ATF.

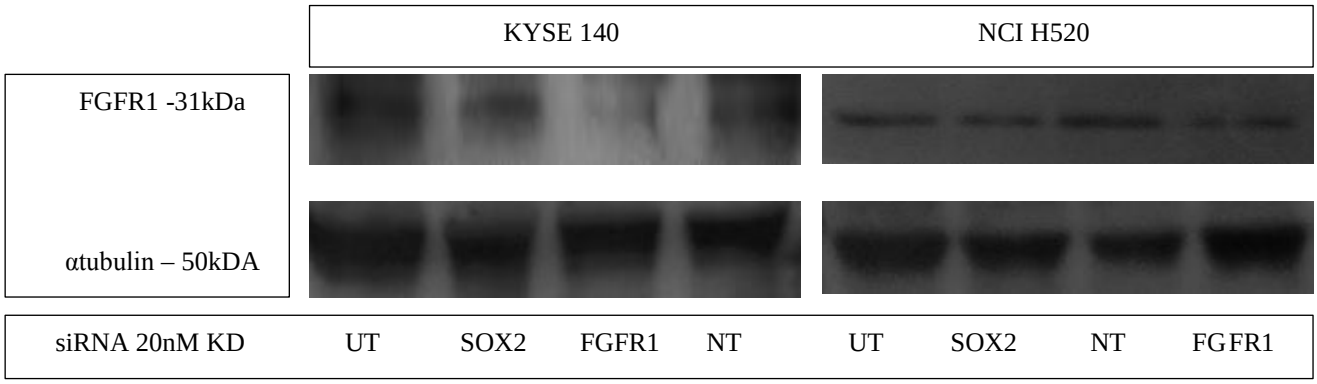


Fig.21. Western blot images showing SOX2 and FGFR1 knockdowns on KYSE and H520 cell stained for FGFR1 **(A)** Comparison of FGFR1 expression levels in KYSE cells with SOX2 and FGFR1 knockdowns **(B)** Comparison of FGFR1 expression levels in H520 cells with SOX2 and FGFR1 knockdowns. **(C-D)** Loading controls stained for αtubulin.

I searched for a ChIP dataset listing the possible binding sites of SOX2 near FGFR1 locus, and identified a SOX2 binding site 8kbs upstream of FGFR1 transcription start site. SOX2 has a binding site near Pax2 gene promoter and EGR3 gene promoter; both transcription factors have binding sites on FGFR1 locus, thus possibly granting the ability for SOX2 to indirectly alter FGFR1 expression by modifying Pax2 and EGR3 expression (Young Lab, Whitehead Institute, Cambridge, Unpublished Data)

A possible pathway between SOX2 and *FGFR1* expression could be further investigated by an *in vitro* model of SOX2 overexpression in normal lung epithelial cells and assessing *FGFR1* expression levels after *SOX2* overexpression. A possible pathway between two oncogenes of SQC would be uncovered if my observation could be replicated and extended. It is known that SOX2 forms a trimeric complex with OCT 3/4 and is capable of initiating expression of many genes critical in embryonic differentiation and pluripotency of esophageal stem cells (Okumura Nakanishi et al. 2005).

If it can be shown that amplified SOX2 may act, at least in part, via FGFR1 induction, this would imply that the available FGFR1 inhibitors may have efficacy on both FGFR1 and SOX2 amplified SQCs.

Future Plans

For a more efficient *in vitro* knockdown of SOX2; an SQC cell line with tetracycline-controlled transcriptional activation built in could be generated that includes a *SOX2* targeting shRNA in its genome, providing a means for high grade *SOX2* inhibition upon doxycycline treatment of cells. An induced model of SOX2 depletion could be more efficient in investigating possible effects of autophagy inhibitors.

To confirm the further decrease in proliferation stemming from a synergy between *FGFR1* knockdown and CQ, evidence of autophagy in FGFR1 depletion must be presented. Also replicating my proliferation experiments would result in a larger data set to work with and might be able to uncover the extent of effect that CQ treatment might have on these cells.

Next move would be to carry out these observations to a model that would resemble the environment of SQC. By carrying out knockdown experiments and autophagy inhibitor treatments, whether by an inducible shRNA method, a stable lentiviral method, or transient siRNA method, on organotypic culture; possible effects of CQ on lung tissue environment can be observed.

During my project plans of co-culturing fibroblast cells embedded on a collagen based gel acting as an intracellular environment with one of my SQC cell lines had been suggested, and this would be the next logical step to carry out knockdown experiment in an *in vitro* model resembling SQC.

Lastly; ChIP-RT- PCR experiments should be conducted to better define my observation of SOX2 altering FGFR1 expression by investigating if SOX2 binds to FGFR1 locus in this system and obtain FGFR1 inhibitors for tests against SOX2 amplified cell lines.

4.2 Interrogation of LKB1 and P16 on SQC via MCC technique

The challenges involved in analyzing small heterogeneous formalin-fixed biopsy samples from preinvasive lesions are significant due to the poor quality and low quantity of DNA available.

Previous analyses of preinvasive lesions have been limited as a result of this. µMCC offers the possibility of precisely measuring copy-number variations in these samples, while at the same time detecting common or rare mutant alleles and allowing a mutant allele frequency.

I have therefore optimized unique primer sets and demonstrated for the first time on a low-grade preinvasive lesion the ability to interrogate two key tumor suppressor genes in parallel for copy number variations and mutations on subnanogram quantities of fragmented FFPE DNA. I have also demonstrated that the MCC technique can be used to interrogate copy number losses on tumor suppressor genes with confidence, alongside investigating copy number amplifications on oncogenes of SQC.

During the course of the project, I have designed multiple sets of copy number primers for both LKB1 and P16 genes. Although these primer sets have been tested only on 3 instances (normal lung DNA, low grade lesion DNA, and high grade lesion DNA), they can be used to interrogate the state of LKB1 and P16 on any number of SQC lesions and samples of other cancer subtypes to understand the state of LKB1/P16 in a greater scale.

During the primer validation process, I had to redesign many copy number and mutation primer sets. This could be due to proximity to genomic regions with a high

melting temperature; a feature that is, in turn, closely correlated with percentage GC content. It was originally considered (Daser et al.2006, McCaughan et al. 2011) that a combination of three aspects of the MCC protocol overcame the issue of PCR efficiency: firstly, the consistent oligonucleotide design parameters; secondly, the hemi-nested 2-phase protocol; and thirdly, a qualitative digital requirement for either a presence or absence of Phase 2 product, meaning that the band intensity or final quantity of each product was not critical.

However, the current data suggest that a property of the template DNA, external to the actual sequence that is being amplified, can strongly influence PCR efficiency in intact DNA. Also the tertiary structure of the DNA on target sequence can strongly influence the efficiency of the primer set.

Interrogations of LKB1 and P16 on low grade dysplasia resulted in the calculation of the relative copy number of 1 for both genes. This implies that the tumor suppressor genes LKB1 and P16 may undergo losses in their respective chromosome regions after development of the lesion onto more severe forms of dysplasia. However my tests include interrogations upon only one lesion and the tests must be performed on multiple sets of SQC low grade lesions to draw a complete picture of the state of LKB1 and P16 in low grade SQC lesion genome. As I have confirmed that the designed copy number markers reflect copy number of target genes with high efficiency, it is only a matter of replicating these experiments on other lesion in terms of future work.

I also achieved to sequence 4 of the 5 highest frequency mutation hotspots on LKB1 gene on low grade SQC dysplasia DNA. While the sequencing results did not indicate the presence of mutations on any of the mutation hotspots, this is the first documented instance of mutation mapping on a low grade SQC lesion using the digitalPCR technique. The primers designed would give more precise and efficient sequencing results than using bulk sequencing techniques, allowing for identification of low frequency mutation events. This ability is based on the two phase amplification technique of the digital PCR procedure, where the bulk amplification takes place in the first phase and second phase allows specific amplification target sequences.

This ability is especially important in sequencing DNA retrieved from FFPE (formalin fixed paraffin embedded) DNA, where the low quantity and fragmented nature of DNA would make any one phase amplification and sequencing method practically impossible. With the mutation primers designed in this project, any frequency of mutation events can be discerned from formalin fixed deteriorated DNA samples.

Future applications of the digital PCR procedure used in this project could include mapping of the mutation hotspots on p16 gene and testing the mutation primers for LKB1 on increased numbers of high grade and low grade lesions, to understand the development into severe dysplasia in the molecular sense.

Acknowledgements

I would like to thank Dr. Frank McCaughan for his excellent supervision, constructive criticism, comments and insight on experimental procedures, for supplying the samples and robotics equipment used in digital PCR, and cell lines used in RNAi screening procedures.

I would also like to thank Hassan Farah for his assistance during experimental procedures and helpful comments and discussion on development of ideas.

I declare that I have personally prepared this report and that it has not in whole or in part been submitted to any other degree or qualification. The work described here is my own, carried out personally unless otherwise stated. All sources of information, including quotations, are acknowledged by means of reference.

5 Bibliography

1. Amaravadi RK, Yu D, Lum JJ, Bui T, Christophorou MA, Evan GI, et al. Autophagy inhibition enhances therapy-induced apoptosis in a Myc-induced model of lymphoma. J Clin Invest 2007; 117: 326–36.
2. Amaravadi, R.K., Lippincott-Schwartz, J., et al. Principles and current strategies for targeting autophagy for cancer treatment. Clin Can Res2011; 17: 654-666.
3. Bass AJ, Watanabe H, Mermel CH, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 2009; 41: 1238–42.
4. Daser A, Thangavelu M, Pannell R, *et al* . Interrogation of genomes by molecular copy-number counting (MCC). *Nat Methods* 2006; **3**: 447–453.
5. Gill RK, Yang SH, Meerzaman D, et al. Frequent homozygous deletion of the LKB1/STK11 gene in non-small cell lung cancer. Oncogene 2011; 30: 3784-3791.
6. Gunja N, Roberts D, McCoubrie D, Lamberth P, Jan A, Simes DC, et al. Survival after massive hydroxychloroquine overdose. Anaesth Intensive Care 2009; 37: 130–3.
7. Hearle N, et al. Frequency and spectrum of cancers in the Peutz–Jeghers syndrome. Clin. Cancer Res 2006; 12: 3209–3215.
8. Hussenet T, Dali S, Exinger J, et al. SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. PLoS One 2010; 5: e8960.
9. Ji H, Ramsey MR, Hayes DN, Fan C, McNamara K, Kozlowski P et al. LKB1 modulates lung cancer differentiation and metastasis. Nature 2007; 448: 807–810.
10. Kamb, A., Gruis, N. A., Weaver-Feldhaus, J., Liu, Q., Harshman, K., Tavtigian, S. V., Stockert, E., Day, R. S., III, Johnson, B. E., Skolnick, M. H. A cell cycle regulator potentially involved in genesis of many tumor types. Science 1994; .264: 436-440.
11. Landis SH, Murray T, Bolden S, Wingo PA.Cancer statistics, 1999.CA Cancer J Clin, 1999; 49: 8e31- 31.
12. McCaughan F, Darai-Ramqvist E, Bankier AT, et al. Microdissection molecular copy-number counting (microMCC) — unlocking cancer archives with digital PCR. *J Pathol* 2008; **216**: 307–316.
13. McCaughan F, Pipinikas CP, Janes SM, et al. Genomic evidence of pre-invasive clonal expansion, dispersal and progression in bronchial dysplasia. J Pathol 2011; 224: 153- 159.

14. McCaughan F, Pole JC, Bankier AT, et al. Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. *Am J Respir Crit Care Med* 2010; **182**: 83–91.
15. Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem 2005; 280, 5307-5317.
16. Que, J., Luo, X., Schwartz, R. J., Hogan, B. L. M. Multiple roles for Sox2 in the developing and adult mouse trachea. Development 2009; 136: 1899-1907.
17. Que, J., Okubo, T., Goldenring, J. R., Nam, K.-T., Kurotani, R., Morrisey, E. E., Taranova, O., Pevny, L. H., Hogan, B. L. M. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 2007;134: 2521-2531.
18. Rabinowitz JD, White E. Autophagy and metabolism. Science 2010; 330:1344–8.
19. Robertson, KD., Jones, PA. Tissue-specific alternative splicing in the human INK4a/ARF cell cycle regulatory locus. Oncogene 1999; 18: 3810-3820,.
20. Romanelli F, Smith KM, Hoven AD. Chloroquine and hydroxychloroquine as inhibitors of human immunodeficiency virus (HIV-1) activity. Curr Pharm Des 2004;10:2643–8.
21. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology 2000; 365-386.
22. Schuck S, Manninen A, et al. Generation of single and double knockdowns in polarized epithelial cells by retrovirus-mediated RNA interference. PNAS 2004; 101:4912-4917.
23. Spiro SG. One hundred years of lung cancer. Am J Respir Crit Care Med 2005; 172: 523-529.
24. Sutherland KD, Berns A. Cell of origin of lung cancer. Mol Oncology 2010; 4: 397-403.
25. Travis WD. Pathology of lung cancer. Clin Chest Med 2002; 23: 65e81. viii.
26. Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. Nat Rev 2010; 10: 116-129.
27. Weiss J, et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Science Trans Med 2010; 2: 62ra93.
28. Wiest, J. S., Franklin, W. A., Otstot, J. T., Forbey, K., Varella-Garcia, M., Rao, K., Drabkin, H., Gemmill, R., Ahrent, S., Sidransky, D., Saccomanno, G., Fountain, J. W., Anderson, M. W. Identification of a novel region of homozygous deletion on chromosome 9p in squamous cell carcinoma of the lung: the location of a putative tumor suppressor gene. Cancer Res1997; 57: 1-6.
29. Wistuba II, Behrens C, Milchgrub S, et al. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. Oncogene 1999; 18: 643–650.