Appendix

Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium:

Rebecca C. Fitzgerald[1], Ayesha Noorani[1], Paul A.W. Edwards[1,2], Nicola Grehan[1], Barbara Nutzinger[1], Caitriona Hughes[1], Elwira Fidziukiewicz[1], Jan Bornschein[1], Shona MacRae[1], Jason Crawte[1], Alex Northrop[1], Gianmarco Contino[1], Xiaodun Li[1], Rachel de la Rue[1], Maria O'Donovan[1,3], Ahmad Miremadi[1,3], Shalini Malhotra[1,3], Monika Tripathi[1,3], Simon Tavaré[2], Andy G. Lynch[2], Matthew Eldridge[2], Maria Secrier[2], Lawrence Bower[2], Ginny Devonshire[2], Juliane Perner[2], Sriganesh Jammula[2], Jim Davies[5], Charles Crichton[5], Nick Carroll[6], Peter Safranek[6], Andrew Hindmarsh[6], Vijayendran Sujendran[6], Stephen J. Hayes[7,14], Yeng Ang[7,8,29], Shaun R. Preston[9], Sarah Oakes[9], Izhar Bagwan[9], Vicki Save[10], Richard J.E. Skipworth[10], Ted R. Hupp[10], J. Robert O'Neill[10,23], Olga Tucker[11,33], Andrew Beggs[11,28], Philippe Taniere[11], Sonia Puig[11],Timothy J. Underwood[12,13], Fergus Noble[12], Jack Owsley[12], Hugh Barr[15], Neil Shepherd[15], Oliver Old[15], Jesper Lagergren[16,25], James Gossage[16,24], Andrew Davies[16,24,], Fuju Chang[16,24], Janine Zylstra[16,24], Ula Mahadeva[16], Vicky Goh[24], Francesca D. Ciccarelli[24], Grant Sanders[17], Richard Berrisford[17], Catherine Harden[17], Mike Lewis[18], Ed Cheong[18], Bhaskar Kumar[18],Simon L Parsons[19], Irshad Soomro[19], Philip Kaye[19], John Saunders[19], Laurence Lovat[20], Rehan Haidry[20], Laszlo Igali[21], Michael Scott[22], Sharmila Sothi[26], Sari Suortamo[26], Suzy Lishman[27], George B. Hanna[31], Krishna Moorthy[31], Christopher J. Peters[31], Anna Grabowska[32], Richard Turkington[34.]

[1] Medical Research Council Cancer Unit, Hutchison/Medical Research Council Research Centre, University of Cambridge, Cambridge, UK
[2] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
[3] Department of Histopathology, Addenbrooke's Hospital, Cambridge, UK
[4]Oxford ComLab, University of Oxford, UK, OX1 2JD
[5]Department of Computer Science, University of Oxford, UK, OX1 3QD
[6]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK, CB2 0QQ
[7]Salford Royal NHS Foundation Trust, Salford, UK, M6 8HD

[8]Wigan and Leigh NHS Foundation Trust, Wigan, Manchester, UK, WN1 2NN
[9]Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK, GU2 7XX
[10]Edinburgh Royal Infirmary, Edinburgh, UK, EH16 4SA
[11]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK, B15 2GW
[12]University Hospital Southampton NHS Foundation Trust, Southampton, UK, SO16 6YD
[13]Cancer Sciences Division, University of Southampton, Southampton, UK, SO17 1BJ
[14]Faculty of Medical and Human Sciences, University of Manchester, UK, M13 9PL
[15]Gloucester Royal Hospital, Gloucester, UK, GL1 3NN
[16] Guy's and St Thomas's NHS Foundation Trust, London, UK, SE1 7EH
[17]Plymouth Hospitals NHS Trust, Plymouth, UK, PL6 8DH
[18]Norfolk and Norwich University Hospital NHS Foundation Trust, Norwich, UK, NR4 7UY
[19]Nottingham University Hospitals NHS Trust, Nottingham, UK, NG7 2UH
[20]University College London, London, UK, WC1E 6BT
[21]Norfolk and Waveney Cellular Pathology Network, Norwich, UK, NR4 7UY
[22]Wythenshawe Hospital, Manchester, UK, M23 9LT
[23]Edinburgh University, Edinburgh, UK, EH8 9YL
[24]King's College London, London, UK, WC2R 2LS
[25]Karolinska Institutet, Stockholm, Sweden, SE-171 77
[26]University Hospitals Coventry and Warwickshire NHS, Trust, Coventry, UK, CV2 2DX
[27]Peterborough Hospitals NHS Trust, Peterborough City Hospital, Peterborough, UK, PE3 9GZ
[28]Institute of Cancer and Genomic sciences, University of Birmingham, B15 2TT
[29]GI science centre, University of Manchester, UK, M13 9PL.
[30]Queen's Medical Centre, University of Nottingham, Nottingham, UK, NG7 2UH
[31]Department of Surgery and Cancer, Imperial College London, UK, W2 1NY
[32]Queen's Medical Centre, University of Nottingham, Nottingham, UK
[33]Heart of England NHS Foundation Trust, Birmingham, UK, B9 5SS.
[34]Centre for Cancer Research and Cell Biology, Queen's University Belfast, Northern Ireland, UK, BT7 1NN.

Supplementary Figure Legends

Supplementary Figure 1. a. Venn diagram showing all EAC cases for which complementary methylation, WGS and RNA-seq data are available. **b.** Same as A but the cases here represent BE. **c.** Principal component analysis on most variable probes computed across all samples including controls. **d.** Cophenetic coefficient computed for estimating optimal rank through NMF. **e.** Average silhouette width for estimating optimal k-means cluster. **f.** Heatmap representing four metagene expression for all samples across four identified subtypes. **g.** Boxplot representing distribution of tumour cellularity computed from whole genome

sequencing data for all samples within individual subtype. **h.** Boxplot representing distribution of mutation burden of all samples within individual subtype. **i.** Boxplot representing distribution of total structural variants of all samples within individual subtype. (p-value: * (<0.05), ** (<0.01), *** (<0.001)).

Supplementary Figure 2. a. Distribution of different sets of clinical categories for all EAC cases across all subtypes. **b.** Age distribution across all subtypes (p-value=0.408). **c.** Distribution of dysplastic status for all BE cases across all subtypes. **d.** Distribution of dysplasia grade (ND: Non Dysplastic, LGD: low grade dysplasia, HGD: high grade dysplasia, IMC: intra mucosal carcinoma) for all BE cases across all subtypes. **e.** Distribution of BE cases with and without adjacent tumour across all subtypes. **f.** Cophenetic coefficient computed from published dataset for estimating optimal rank through NMF. **g.** Average silhouette width from published dataset for estimating optimal k-means cluster. **h.** Heatmap comparing methylation patterns between our dataset (left) and published dataset (right).

Supplementary Figure 3. Barchart representing driver genes with proportion of cases being altered in individual subtype separated by BE and EAC (* p-value < 0.05).

Supplementary Figure 4. Correlation plot comparing expression of target transcription factor and median level of methylation estimated from all distal probes which showed positive enrichment for that particular transcription factor.

Supplementary Figure 5. a. List of most enriched pathways in Subtype II computed through gene set enrichment analysis. **b.** List of most enriched pathways in Subtype III computed through gene set enrichment analysis. **c.** List of most enriched pathways in Subtype IV computed through gene set enrichment analysis **d.** IHC staining for Granzyme B on four different chemo-naïve cases representative of four subtypes.

Supplementary Figure 6. a. Gene ontology showing enrichment of key biological process when analysed on silenced genes. **b.** Pathways enriched on silenced genes computed through Ingenuity. **c.** Correlation between methylation and expression for MGMT gene across all cases from four different subtypes including controls. **d.** Same as C for CHFR gene. **e.** Expression levels of MGMT in different organoids arranged in order of low to high represented as heatmap. **f.** Expression levels of CHFR in different organoids arranged in order of low to high represented as heatmap. Table below shows response of selected organoids (highlighted in bold and with same colour as represented in heatmap) toward docetaxel treatment.

Supplementary Figure 7. a. Principal component analysis showing organization of samples from different tissue types on the basis of their methylation profiles. **b.** Principal component analysis showing organization of samples from different tissue types on the basis of their expression profiles. **c.** Gene ontology analysis showing top 30 biology process enrichment in squamous, gastric and duodenum. **d.** Principal component analysis showing similarity in methylation profiles between samples from varied tissue types derived from different studies. **e.** Same samples as shown in **d**, grouped on the basis of hierarchical clustering of their methylation profiles. **f.** Boxplot showing distribution of mutation burden of BE cases across all four subtypes. **g.** Boxplot showing proportion of genome altered through copy number aberrations for BE cases across all four subtypes.

Supplementary Table 1: Clinical annotation of EAC cases

Supplementary Table 2: Clinical annotation of BE cases

Supplementary Table 3: List of epigenetically silenced genes

Supplementary Table 4-5: Gene Ontology and pathway analysis on epigenetically silenced genes

Supplementary Table 6-9: GSEA based pathway enrichment in Subtype I, II, III and IV

Supplementary Methods

Gene Expression Profiling

Total RNA was extracted using All Prep DNA/RNA kit from Qiagen and the quality was checked on an Agilent 2100 Bioanalyzer using the RNA 6000 nano kit (Agilent). The Qubit high sensitivity RNA assay kit from Thermo Fisher was used for quantification. Libraries were prepared from 250ng RNA, using TruSeq Stranded Total RNA Library Prep Gold (Ribo-zero) kit and ribosomal RNA (nuclear, cytoplasmic and mitochondrial rRNA) was depleted, whereby biotinylated probes selectively bind to ribosomal RNA molecules forming probe-rRNA hybrids. These hybrids were pulled down using magnetic beads and rRNA depleted total RNA was reverse transcribed. The libraries were prepared according to the Illumina protocol[1]. Paired-end 75bp sequencing on HiSeq4000 generated the paired end reads.

Histone Modification Profiling

We downloaded ChIP-seq data for H3K27me3 and H3K27ac histone modifications in aligned form for esophagus, gastric and duodenum tissues from ENCODE[2, 3]/ROADMAP[4] consortium data portal. We quantified two histone modifications across all normal tissues on extended regions (5kb both up and downstream from centre) that were gaining methylation in Subtype I as compared to controls. The quantification shown in Figure 2 was generated using ngs.plot[5].

Organoid culture and Drug treatment

The primary organoid cultures were derived from one normal gastric case and eight EAC cases included in the OCCAMS/ICGC sequencing study. Detailed organoid culture and derivation method have been previously described[6]. Regarding the drug treatment, the seeding density for each line was optimised to ensure cell growth in the logarithmic growth phase. Cells were seeded in complete medium for 24 hours then treated with compounds at a 9-point half-log serial dilutions for 6 days. Cell viability was assessed using CellTiter-Glo (Promega) after drug incubation. The concentrations of a compound causing 50% growth inhibition relative to the vehicle control (GI50) were determined by nonlinear regression dose-response analysis and the area under the curve (AUC) was calculated using GraphPad Prism.

Granzyme B (GZMB) Immunohistochemistry

FFPE tissues are sectioned at 4um thickness, floated onto charged glass slides and dried at 37ºC overnight. Deparaffinisation (69°C for 32min), antigen retrieval (pH8 with CC1), peroxidase inhibition (Discovery inhibitor/inhibitor CM) and indirect IHC are conducted with the automated Ventana Discovery Ultra platform and Leica Bond. Primary antibody for Granzyme B from Abcam (EPR8260) was used in this study.

Performance of primary antibodies is compared against negative isotype controls. Mouse monoclonal antibodies are ready-to-use (RTU) preparations manufactured by Roche. Rabbit primary Ab negative control is produced by DAKO at a concentration of 1500 ul/ml. This negative isotype is diluted to match the concentration of the tested primary antibody. For counterstaining and post-counterstaining, haematoxylin and bluing reagent are consecutively applied to the sections, and each is incubated for 16 minutes. Slides are washed with reaction buffer after each incubation, throughout the automated process.

Stained slides are dehydrated and automatically cover-slipped using the Leica Autostainer ST020. They are digitally scanned by Aperio Scanscope XT at a 20X resolution. Images are annotated digitally using the HALO® TM digital image analysis software v2.1.1637.11 (Indica Labs, Corrales, NM, USA). In total, IHC data for GZMB was available for 17 EAC cases.

MGMT Immunohistochemistry

Paraffin embedded sections of 3.5 µm were used for immunohistochemistry by a Bond Max autostainer according to the manufacturer's instruction (Leica Microsystems). Primary antibodies MGMT (MT3.1, Merck, 1:100 dilution) were optimized (incubation 30 mins) and applied with controls.

Pathway Analysis

Gene Ontology and Pathway analysis of silenced gene was performed using David[7] and IPA (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis).

References

1. Nagai K, Kohno K, Chiba M, et al. Differential expression profiles of sense and antisense transcripts between HCV-associated hepatocellular carcinoma and corresponding non-cancerous liver tissue. Int J Oncol 2012;40:1813-20.
2. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.
3. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 2018;46:D794-D801.
4. Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317-30.
5. Shen L, Shao N, Liu X, et al. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 2014;15:284.
6. Li X, Francies HE, Secrier M, et al. Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics. Nat Commun 2018;9:2983.
7. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.