Somatic Cancer Genetics in the UK: Results from Phase One of the Cancer Research UK Stratified Medicine Programme

Lindsay CR, Shaw EC, Blackhall F , Blyth KG, Brenton JD, Chaturvedi A, Clarke N, Dick C, Evans TRJ, Hall G, Hanby AM , Harrison DJ, Johnston S, Mason MD, Morton D, Newton-Bishop J, Nicholson AG, Oien KA , Popat S, Rassl D, Sharpe R, Taniere P, Walker I, Wallace WA, West NP, Butler R, Gonzales de Castro D, Griffiths M, Johnson PWM.

Supplementary Methods and Data

Supplementary Methods

The Cancer Research UK Stratified Medicine Programme

Operational delivery of the programme was monitored through monthly collection of data from the operational lead at each clinical hub or technical hub in the form of a completed Microsoft Excel spreadsheet template containing details of aggregated patient accrual and numbers of samples at each stage of the process, as well as detailed sample level data on test failures from the technical hubs.

Analysis, Interpretation and Reporting

Laboratories performing the genetic analysis SMP1 samples were permitted to set sample requirements and use techniques they already had in use (Supplementary Methods Table 1), but demonstrate reproducibility of the results through participation in a bespoke external quality assessment (EQA) scheme run by the UK National External Quality Assessment Service (NEQAS) for molecular genetics. FFPE tissue sections and a peripheral blood sample from each patient were forwarded from each of the eight clinical hubs to one of the three technology hubs for genetic analysis. Analysis of sequence changes, particularly in tumour suppressor genes, involved online databases and literature searches to try and establish pathogenicity and interpretation of low level variants, especially in samples with high signal-to-noise ratio due to poor quality DNA and/or formalin-induced DNA damage. Close attention was also applied to *TMPRSS2-ERG* analysis by fluorescent in situ hybridisation (FISH) in prostate cancer, due to the frequent occurrence of complex rearrangements associated with copy number aberrations in the partner genes.

Each technology hub received sample requests as electronic messages in open XML (eXtensible Markup Language) format, transferred via a dedicated CRUK-hosted sFTP (secure file transfer

protocol) server and NHS approved client software; generated by a staff member at the clinical hub to notify them that a blood or tissue sample had been dispatched in the post. The XML schema specification defined standard contents of the XML, including the sample study identifier, sample type and fields for the return of results when available. This message was retrieved by the TH from the sFTP server and a similar message was returned to the CH to notify them that the sample had been received by the TH. Once molecular analysis was complete, staff at the TH conveyed the molecular report as an XML message to the relevant CH. For each TH, the CRUK sFTP and XML messaging was linked to the local TH database so that patient data could be directly exported to the sFTP without manual transcription. Results were transmitted electronically to clinical centres for inclusion in individual patient medical records, but accompanied by a statement that 'these results are intended for research purposes'.

A clinical dataset including diagnostic, treatment and outcome data was compiled at clinical sites for participating patients and submitted to the lead cancer registry for England. This clinical dataset was based on existing NHS information standards, including attributes drawn from the enhanced cancer registration dataset for England, the Cancer Outcomes and Services Dataset (COSD), with data item definitions from the NHS Data Dictionary and use of accepted coding systems such as ICD-10 (10^{th} revision of the World Health Organization International Statistical Classification of Diseases and Related Health Problems) and SNOMED RT (Systematized Nomenclature Of Medicine Clinical Terms Reference Terminology) morphology codes.

Death registration data, including the date and cause of death for patients who died during the course of the study, was requested at six monthly intervals by the National Cancer Registration Service from the Office of National Statistics for patients in England and Wales and as a one-off request in 2015 from NHS National Services Scotland for patients in Scotland.

Following removal of patient-identifiable demographic data items, an extract of collated clinical, pathological, treatment and outcomes data for SMP patients was sent to the Department of

Computing Sciences at the University of Oxford and incorporated into a single Microsoft Access database. This database was composed of a number of linked tables, designed and developed according to a relational data model.

Database Analysis

Manual data curation and cleansing was performed on an extract obtained from the central research database, containing a subset of data items exported into a Microsoft Excel spreadsheet. This data cleansing involved creation of six disease-based cohorts, rationalising SNOMED codes from the diagnosis and pathology tables where they differed, leaving one SNOMED code per row that provided the most diagnostic detail (e.g. leaving 'M81403, adenocarcinoma' in place of 'M80103, carcinoma not otherwise specified' where both were present).

The format of the genetic data was simplified where possible to facilitate aggregated data analysis. Since the full reports included both the sequence/ coding and predicted protein change, this involved reducing it to the predicted protein change only where it was possible to determine this (e.g. for an *EGFR* result reported as 'c.2573T>G (p.L858R)' this was simplified to 'L858R'). Categorisation of the clinical data by tumour stage and chemotherapy regimen were also undertaken to rationalise these heterogeneous data and simplify analysis. Data completeness varied between the clinical sites and between data items (Table 4). There was a degree of overlap and redundancy between data items, allowing some gaps to be filled: for example the tumour stage could be determined, if the individual components of TNM (tumour/nodes/metastasis) classification were available.

Although there was an aspiration at the outset of the programme for the dataset to be populated by automated data extraction from electronic patient records, informal feedback during the

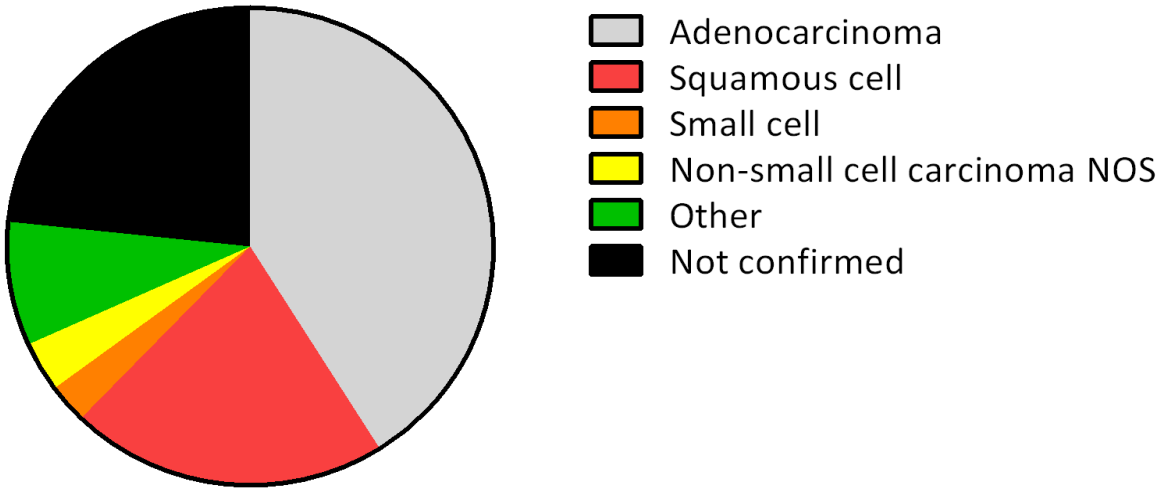
programme indicated that, due to a lack of standardisation across NHS systems, there was a broad requirement for manual compilation of data from multiple different clinical systems.

Supplementary Methods Table 1. Genetic regions of interest and techniques used for analysis during SMP1

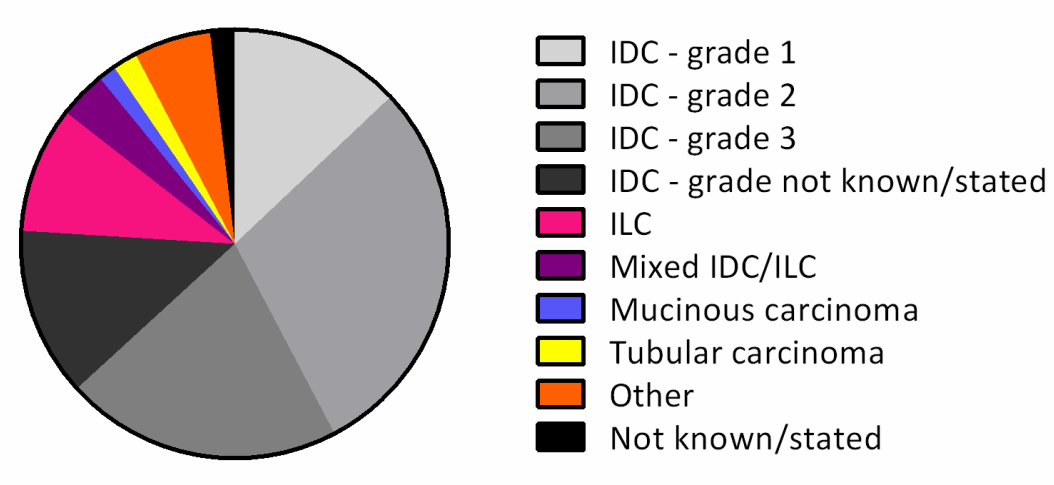
Gene	Scope of test	Lab 1	Lab 2	Lab 3
BRAF	Initially codons 599-601 in exon 15, then exons 11 and 15 from September 2012	Qiagen TheraScreen Pyromark BRAF Assay, pyrosequencing, Roche Cobas® 4800 system, direct sequencing	Pyrosequencing	CE-SSCA +/- sequencing
		Exon 11 sequencing		
DDR2	Exons 3-18; added in September 2012	Sequencing		
EGFR	Exons 18-21	Qiagen TheraScreen Pyromark EGFR Assay	Pyrosequencing or fragment length analysis	CE-SSCA
EML4-ALK	Confirm presence of breakpoint in *ALK* gene	FISH, break apart probe		
KIT	Exons 11, 13, 17	Sequencing		CE-SSCA
KRAS	Codons 12, 13, 61, 146	Qiagen TheraScreen Pyromark/ Rotor-Gene Q PCR KRAS Assay	Pyrosequencing	Cobas® 4800
NRAS	Codons 12, 13, 61	Qiagen TheraScreen Pyromark NRAS Assay		CE-SSCA
PIK3CA	Exons 9, 20	Pyrosequencing, Snapshot or Qiagen ARMS kit		
PTEN	Exons 2-10	Sequencing	HRM analysis, sequencing	CE-SSCA
TMPRSS2- ERG	Confirm presence of rearrangement	FISH (triple colour probe)		
TP53	Exons 4-9	Sequencing		CE-SSCA

ARMS: Amplification Refractory Mutation System; CE-SSCA: capillary electrophoresis-single strand conformation analysis; FISH: fluorescent in situ hybridisation; HRM: high resolution melt; STR: short tandem repeat.

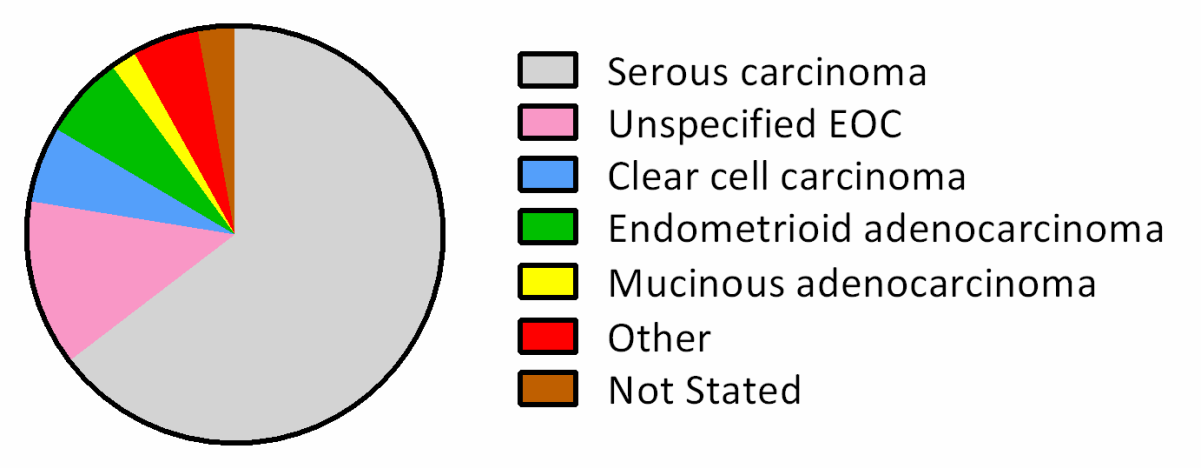
Supplementary Results - Figures



Supplementary Figure 1. Overview of lung cancer histological subtypes represented in the SMP1 cohort. 'Other' refers to a mix of rarer pathologies including carcinoid, large cell carcinoma, and salivary-type carcinomas. NOS = not otherwise specified.



Supplementary Figure 2. Overview of breast cancer histological subtypes represented in SMP1 cohort. IDC= invasive ductal carcinoma, ILC = invasive lobular carcinoma.



Supplementary Figure 3. Overview of ovarian cancer histological subtypes represented in SMP1 cohort.

Supplementary Results - Tables

Supplementary Table 1 Baseline characteristics of patients with lung cancer

Characteristic	Number	%
Gender		
Male	975	51.7
Female	866	45.9
Not Stated	44	2.4
Age, years		
Median	68	
Range	20-88	
Ethnicity		
White British/Irish	1280	67.9
Other	127	6.7
Not Stated	478	25.4
Histological Subtype		
Adenocarcinoma	774	41.1
Squamous Cell	399	21.2
Small Cell	50	2.6
NSCLC NOS	64	3.4
Other	158	8.4
Not Stated	440	23.3
Stage		
I-II	1006	53.4
III	396	21.0
IV	365	19.4
Not Stated	118	6.2

Abbreviations: NSCLC NOS = non-small cell lung cancer not otherwise specified. RMH = Royal Marsden Hospital.

Supplementary Table 2 Prevalence of lung gene panel mutations/modifications (*KRAS, EGFR, ALK, BRAF*) in the SMP1 lung cancer patient cohort – failed tests included

	Patient Numbers (%)			
Gene (including combinations)	**Adenocarcinoma**	**Squamous Cell**	**Small Cell**	**NSCLC NOS**
Total samples	774/1885 (41.1)	399/1885 (21.2)	48/1885 (2.55)	64/1885 (3.4)
EGFR	92/774 (11.89)	3/399 (0.75)	0/48 (0)	4/64 (6.25)
KRAS	287/774 (37.08)	12/399 (3.01)	0/48 (0)	14/64 (21.88)
ALK	19/774 (2.45)	2/399 (0.5)	0/48 (0)	2/64 (3.13)
DDR2	x	12/175 (6.86)	0/4 (0)	1/6 (16.67)
***BRAF* (exon 15)**	11/774 (1.42)	3/399 (0.75)	0/48 (0)	0/64 (0)
***BRAF* (exon 11)**	7/343 (2.04)	2/189 (1.06)	0/14 (0)	0/16 (0)

DDR2 and *BRAF* exon 11 testing was introduced midway through SMP1, accounting for the smaller total sample number in these rows. Failed tests were included in total samples to calculate prevalence in this analysis. *DDR2* testing was performed for squamous cell carcinoma samples only. NSCLC NOS = non-small cell lung cancer not otherwise specified, i.e. histological subtype could not be further sub-classified.

Supplementary Table 3 Prevalence of lung gene panel mutations/modifications (*KRAS, EGFR, ALK, BRAF*) in the SMP1 lung cancer patient cohort – failed tests excluded

	Patient Numbers (%)			
Gene (including combinations)	**Adenocarcinoma**	**Squamous Cell**	**Small Cell**	**NSCLC NOS**
Total samples	774/1885 (41.1)	399/1885 (21.2)	48/1885 (2.55)	64/1885 (3.4)
EGFR	92/711 (12.94)	3/365 (0.82)	0/47 (0)	4/59 (6.78)
KRAS	287/647 (44.36)	12/304 (3.95)	0/42 (0)	14/55 (25.45)
ALK	19/774 (2.45)	2/399 (0.5)	0/48 (0)	2/62 (3.23)
DDR2	x	12/119 (10.08)	0/3 (0)	1/4 (25)
***BRAF* (exon 15)**	11/712 (1.54)	3/383 (0.78)	0/46 (0)	0/58 (0)
***BRAF* (exon 11)**	7/318 (2.2)	2/169 (1.18)	0/14 (0)	0/16 (0)

DDR2 and BRAF exon 11 testing was introduced midway through SMP1, accounting for the smaller total sample number in these rows. Failed tests were excluded from total samples to calculate prevalence in this analysis. DDR2 testing not performed for adenocarcinoma cases. NSCLC NOS = non-small cell lung cancer not otherwise specified, i.e. histological subtype could not be further sub-classified.

Supplementary Table 4 Prevalence of *EGFR* mutation type in the SMP1 lung cancer cohort

***EGFR* Mutation Type**	**Mutation numbers (%)**
Sensitising	67/92 (72.83)
Resistance	10/92 (10.87)
Sensitising and Resistance	3/92 (3.26)
Unknown	12/92 (13.04)

'Sensitising' refers to L858R point mutations and exon 19 deletions, 'resistance' refers to exon 20 insertions and T790M.

Supplementary Table 5 Baseline characteristics of patients with breast cancer who participated in Phase I of the Stratified Medicine Programme

Characteristic	Number	%
Sex		
Female	1858	99.2
Male	15	0.8
Age, years		
Median	61	
Range	27-96	
Ethnicity		
White British	1215	64.9
Other	124	6.6
Not Stated	534	28.5
Histological Subtype		
Invasive Ductal Carcinoma (IDC)	1423	76
Invasive Lobular Carcinoma (ILC)	179	9.6
Mixed IDC/ILC	67	3.6
Mucinous adenocarcinoma	25	1.3
Tubular adenocarcinoma	35	1.9
Other	109	5.8
Not Stated	35	1.8
Stage		
I-II	1455	77.7
III	162	8.6
IV	48	2.6
Not Stated	208	11.1
Grade		
I	319	17
II	738	39.4
III	492	26.3
Not Stated	324	17.3
Lymphovascular Invasion		
Present	466	24.9
Not Present	1118	59.7
Not Stated	289	15.4
ER status		
Positive	751	40.1
Negative	115	6.1
Not Stated	1007	53.8
PR status		
Positive	271	14.5
Negative	136	7.3
Not Stated	1466	78.2
HER2 status		
Positive	142	7.6
Negative	665	35.5
Not Stated	1066	56.9
Triple negative		
ER- PR- HER2-	60	3.2
ER- PR unknown HER2-	70	3.7

Supplementary Table 6 Prevalence of breast gene panel mutations (*BRAF, PTEN, PIK3CA, TP53*) in patients enrolled in SMP1 diagnosed with breast cancer - failed tests included

	Patient Numbers (%)				
Gene (including combinations)	**IDC**	**ILC**	**Mixed IDC/ILC**	**Mucinous**	**Tubular**
Total samples	1423/1873 (76)	179/1873 (9.6)	67/1873 (3.6)	25/1873 (1.3)	35/1873 (1.9)
BRAF	0/1055 (0)	0/131 (0)	0/44 (0)	0/18 (0)	0/25 (0)
PIK3CA	420/1423 (29.5)	60/179 (33.5)	18/67 (26.9)	3/25 (12)	17/35 (48.6)
PTEN	65/1423 (4.6)	10/179 (5.6)	5/67 (7.5)	0/25 (0)	1/35 (2.9)
TP53	389/1423 (27.3)	6/179 (3.4)	4/67 (6)	4/25 (16)	0/35 (0)

BRAF testing was introduced midway through SMP1, accounting for the smaller total sample number in these rows. Failed tests were included in this analysis. IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma.

Supplementary Table 7 Prevalence of breast gene panel mutations (*BRAF, PTEN, PIK3CA, TP53*) in patients enrolled in SMP1 diagnosed with breast cancer - failed tests excluded

	Patient Numbers (%)				
Gene (inc combinations)	**IDC**	**ILC**	**Mixed IDC/ILC**	**Mucinous**	**Tubular**
Total samples	1423/1873 (76)	179/1873 (9.6)	66/1873 (3.5)	25/1873 (1.3)	35/1873(1.9)
BRAF	0/977 (0)	0/122 (0)	0/37 (0)	0/18 (0)	0/23 (0)
PIK3CA	420/1299 (32.3)	60/167 (35.9)	18/59 (30.5)	3/25 (12)	17/31 (54.8)
PTEN	65/1069 (6.1)	10/131 (7.6)	5/48 (10.4)	0/22 (0)	1/30 (3.3)
TP53	389/1152 (33.8)	6/134 (4.5)	4/44 (9.1)	4/21 (19.1)	0/30 (0)

BRAF testing was introduced midway through SMP1, accounting for the smaller total sample number. Failed tests were not included in this analysis. IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma.

Supplementary Table 8 Baseline characteristics of patients in the SMP1		
Characteristic	**Number**	**%**
Gender		
Male	965	60.1
Female	619	38.6
Not Stated	21	1.3
Age, years		
Median	69	
Range	24-97	
Ethnicity		
White British	1071	66.7
Other	95	5.9
Not Stated	439	27.4
Histological Subtype		
Adenocarcinoma (excluding subtypes)	1508	94.0
Mucinous adenocarcinoma	52	3.2
Other	28	1.7
Not Stated	17	1.1
TNM Stage (5th edition)		
I-II	579	36.1
III	602	37.5
IV	241	15.0
Not Stated	183	11.4
Differentiation/Grade		
Well	114	7.1
Moderate	1126	70.2
Poor	178	11.1
Not Stated	187	11.6
Lymphovascular Invasion		
Present	539	33.6
Not Present	560	34.9
Not Stated	506	31.5

Supplementary Table 9 Prevalence of colorectal gene panel mutations (*BRAF, KRAS, NRAS, PIK3CA, TP53*) in patients in the SMP1 CRC cohort - failed tests included			
	Patient Numbers (%)		
Gene (including combinations)	**Non-mucinous adenocarcinoma**	**Mucinous carcinoma**	**P**
Total samples	1508/1605 (94)	52/1605 (3.2)	
BRAF	144/1508 (9.5)	20/52 (38.5)	<0.0001*
KRAS	581/1508 (38.5)	23/52 (44.2)	0.469
NRAS	61/1508 (4.0)	0/52 (0)	0.264
TP53	824/1508 (54.6)	3/52 (5.8)	<0.0001*
PIK3CA	158/1508 (10.5)	11/52 (21.2)	0.022*
Differences in incidence of each mutation between non-mucinous and mucinous adenocarcinoma subtypes were assessed. Failed tests were included in this analysis. p values obtained from Fisher's exact test.			

Supplementary Table 10 Prevalence of colorectal gene panel mutations (BRAF, *KRAS, NRAS, PIK3CA, TP53*) in the SMP1 CRC cohort - failed tests excluded		
	Patient Numbers (%)	
Gene (including combinations)	**Adenocarcinoma**	**Mucinous carcinoma**
Total samples	1508/1605 (94)	52/1605 (3.2)
BRAF	144/1360 (10.6)	20/50 (40)
KRAS	581/1318 (44.1)	23/49 (46.9)
NRAS	61/1392 (4.4)	0/48 (0)
TP53	824/1217 (67.7)	3/45 (6.7)
PIK3CA	158/1398 (11.3)	11/36 (30.6)
Failed tests were not included in this analysis. p values obtained from Fisher's exact test.		

Supplementary Table 11 Baseline characteristics of SMP1 prostate cancer		
Characteristic	**Number**	**%**
Patient Numbers	1359	100
Age, years		
Median	65	
Range	35-89	
Ethnicity		
White British	730	53.7
Other	92	6.8
Not Stated	537	39.5
Histological Subtype		
Adenocarcinoma	1247	91.8
Other	4	0.3
Not Stated	108	7.9
Stage		
I-II	430	31.6
III	238	17.5
IV	78	5.7
Not Stated	613	45.2
Lymphovascular Invasion		
Present	58	4.3
Not Present	374	27.5
Not Stated	927	68.2

Supplementary Table 12 Prevalence of prostate cancer gene panel mutations/rearrangements (*BRAF, PTEN, TMPRSS2-ERG*) in patients enrolled in SMP1 diagnosed with prostate cancer		
	Patient Numbers (%)	
Gene (inc combinations)	**Adenocarcinoma (failed tests included)**	**Adenocarcinoma (failed tests excluded)**
Total samples	1247/1359 (91.8)	
TMPRSS2-ERG	501/1247 (40.2)	501/1117 (44.9)
PTEN	67/1247 (5.4)	67/944 (7.1)
BRAF	11/937 (1.2)	11/898 (1.2)
As *BRAF* testing in prostate cancer was introduced after commencing SMP1, fewer samples were tested for this gene.		

Supplementary Table 13 Baseline characteristics of SMP1 ovarian cancer patient

Characteristic	Number	%
Patient Numbers	557	100
Age, years		
Median	64	
Range	20-89	
Ethnicity		
White British	358	64.3
Other	33	5.9
Not Stated	166	29.8
Histological Subtype		
Serous	360	64.6
Clear Cell	33	5.9
Endometrioid	36	6.5
Mucinous	11	2.0
EOC, unspecified	72	12.9
Other	29	5.2
Not Stated	16	2.9
FIGO Stage		
I	60	10.8
II	57	10.2
III	268	48.1
IV	82	14.7
Not Stated	90	16.2

EOC=epithelial ovarian cancer.

Supplementary Table 14 Prevalence of ovarian gene panel mutations (*PTEN, PIK3CA, TP53, BRAF*) between histological subtypes in SMP1 ovarian cancer patient cohort - failed tests included

Gene (including combinations)	All	Serous	Clear Cell	Endometrioid	Mucinous	EOC, unspecified
Total samples	557/557 (100)	360/557 (64.6)	33/557 (5.9)	36/557 (6.5)	11/557 (2.0)	72/557(12.9)
BRAF	12/516 (2.3)	7/327 (2.1)	0/32 (0)	0/34 (0)	0/11 (0)	0/67 (0)
PIK3CA	35/557 (6.3)	6/360 (1.7)	9/33 (27.3)	9/36 (25.0)	0/11 (0)	8/72 (11.1)
PTEN	23/557 (4.1)	2/360 (0.6)	4/33 (12.1)	7/36 (19.4)	0/11 (0)	9/72 (12.5)
TP53	265/557 (47.6)	181/360 (50.3)	7/33 (21.2)	12/36 (33.3)	5/11 (45.5)	38/72 (52.8)

BRAF testing was introduced midway through SMP1, accounting for the smaller total sample number in these rows. Failed tests included in this analysis. EOC=epithelial ovarian cancer.

Supplementary Table 15 Prevalence of ovarian gene panel mutations (*PTEN, PIK3CA, TP53, BRAF*) between histological subtypes in SMP1 ovarian cancer patient cohort - failed tests excluded

Gene (including combinations)	All	Serous	Clear Cell	Endometrioid	Mucinous	EOC, unspecified
Total samples	557/557 (100)	360/557 (64.6)	33/557 (5.9)	36/557 (6.5)	11/557 (2.0)	72/557(12.9)
BRAF	12/482 (2.5)	7/300 (2.3)	0/30 (0)	0/32 (0)	0/11 (0)	0/66 (0)
PIK3CA	35/469 (7.5)	6/295 (2.0)	9/29 (31.0)	9/26 (34.6)	0/10 (0)	8/67 (11.9)
PTEN	23/378 (6.1)	2/246 (0.8)	4/22 (18.2)	7/27 (25.9)	0/8 (0)	9/44 (20.5)
TP53	265/430 (61.6)	181/277 (65.3)	7/24 (29.2)	12/28 (42.9)	5/9 (55.6)	38/58 (65.5)

BRAF testing was introduced midway through SMP1, accounting for the smaller total sample number in these rows. Failed tests were not included in this analysis. EOC=epithelial ovarian cancer.

Supplementary Table 16 Baseline characteristics of SMP1 malignant melanoma cohort

Characteristic	Number	%
Sex		
Male	271	50.7
Female	241	45.0
Not Stated	23	4.3
Age, years		
Median	61.5	
Range	18-96	
Ethnicity		
White British	416	77.8
Other	18	3.4
Not Stated	101	18.9
Clinical/Histological Origin		
Superficial Spreading	86	16.1
Nodular	50	9.3
Melanoma NOS	328	61.3
Other	27	5.0
Not Stated	44	8.2
Lymphovascular Invasion		
Present	50	9.3
Not Present	47	8.8
Not Stated	438	81.9

NOS=not otherwise specified.

Supplementary Table 17 Prevalence of melanoma gene panel mutations in SMP1

	Patient numbers (%)	
Gene (including	**Failed tests included**	**Failed tests excluded**
Total samples	535/535 (100)	535/535 (100)
NRAS	124/535 (23.2)	124/419 (29.6)
PIK3CA	8/535 (1.5)	8/399 (2)
BRAF	232/535 (43.4)	232/508 (45.7)
KIT	7/535 (1.3)	7/338 (2.1)

NOS=not otherwise specified.