Supplementary Material

1. BriTROC investigators and Trial Management Group

Sites, investigators and numbers recruited

Beatson West of Scotland Cancer Centre, Glasgow – Rosalind M Glasspool, Iain McNeish, 52
Addenbrooke's Hospital, Cambridge – James D Brenton, 29
St Bartholomew's Hospital, London – Michelle Lockley, Elly Brockbank, 28
Hammersmith Hospital, London – Hani Gabra, Christina Fotopoulou, 27
Mount Vernon Cancer Centre, Northwood – Marcia Hall, 20
Western General Hospital, Edinburgh – Charlie Gourley, 16
St James University Hospital, Leeds – Geoff Hall, 11
The Christie Hospital, Manchester – Andrew Clamp, 11
Guy's and St Thomas' Hospitals, London – Ana Montes, 6
St Mary's Hospital, Manchester – Richard Edmondson, 6
Bristol Haematology and Oncology Centre, Bristol – Axel Walther, 5
City Hospital, Birmingham – Sudha Sundar, 4
Queen Elizabeth Hospital, Gateshead – Raj Naik, 3
Belfast City Hospital – Richard Kennedy, 2

Lead pathologists

Queen Elizabeth University Hospital, Glasgow – David Millan
Addenbrooke's Hospital, Cambridge – Mercedes Jimenez-Linan

Trial Management Group

Iain McNeish (co-Chief Investigator), James Brenton (co-Chief Investigator) Liz-Anne Lewsley (Project Manager), James Paul (Study Statistician), Hani Gabra (Investigator), Darren Ennis (Translational Research Scientist), Cheryl Wilson (study co-ordinator), Paul Dearie (Sponsor Representative).

2. Full inclusion and exclusion criteria

Inclusion

1. Patients with recurrent histologically-proven ovarian cancer, primary peritoneal carcinoma or fallopian tube cancer of high grade serous and high grade endometrioid subtypes. Patients who have a diagnosis of ovarian cancer with a known germline mutation in *BRCA1* or *BRCA2* will also be eligible for inclusion regardless of histological subtype. Patients who are having a diagnostic image-guided biopsy maybe consented and study biopsy taken while awaiting pathological review. Eligible patients who have had samples collected under generic research consent may be registered retrospectively only after full discussion between the site, Chief Investigator and the co-ordinating trials unit (and BriTROC-1 specific consent obtained).
2. Patients must have received at least one line of platinum-containing chemotherapy
3. Availability of formalin-fixed, paraffin-embedded tissue taken at the time of original diagnosis of high grade serous ovarian cancer. This may be primary surgical debulking specimen OR core biopsy. For those with only a core biopsy from time of diagnosis, availability of specimen taken at interval debulking surgery is desirable, but not essential.
4. Patients must have disease deemed suitable for imaging-guided biopsy (ultrasound or CT) by an experienced radiologist or suitable for intra-operative biopsy during secondary debulking surgery as determined by an experienced gynaecological oncology surgeon. Other biopsies, such as skin deposits, are also acceptable. However, this must be confirmed with the Cancer Research UK Clinical Trials Unit prior to patient registration.
5. Age ≥ 18 years.
6. Written informed consent.
7. Able to apply with study procedures.
8. Life expectancy > 3 months
9. No contraindication to biopsy as appropriate.

Exclusion

1. Ovarian, primary peritoneal or fallopian tube cancer of low grade serous, grades 1 or 2 endometrioid, clear cell or carcinosarcoma/malignant mixed mesodermal (MMMT) subtypes unless associated with known germline mutation in *BRCA1* or *BRCA2*.
2. Borderline/low malignant potential tumours
3. Any non-epithelial ovarian malignancy
4. Patients with asymptomatic rising CA125 with no radiological evidence of recurrent ovarian cancer.
5. Original diagnosis of high grade serous cancer made on cytology only

3. Ineligibility and declining to participate

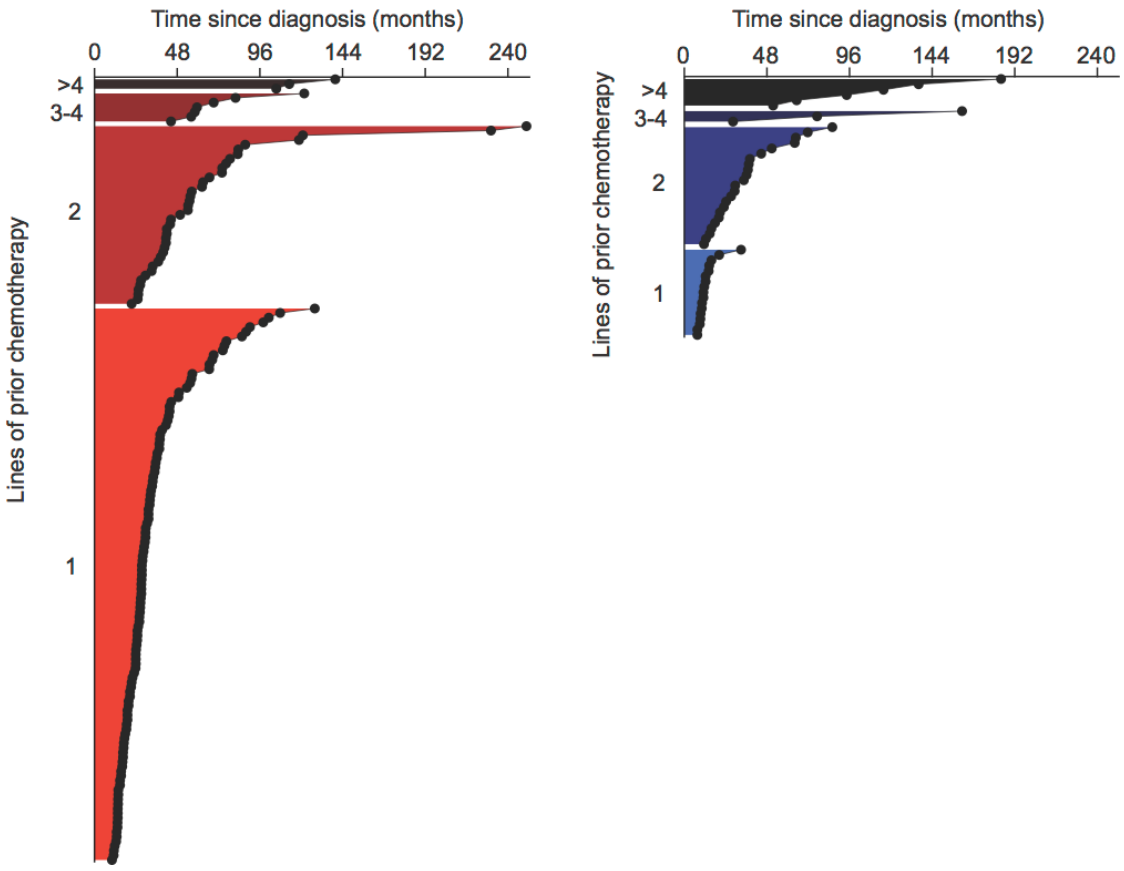
BriTROC-1 Screening Logs	
1. Patient ineligible	182
2. Patient declined as unhappy with proposed trial	38
3. Patient declined for other reason	79
TOTAL	**299**
Reasons for ineligibility	
Is the patient over 18 years of age?	0
Has the patient given written informed consent?	0
Does the patient have recurrent histologically proven ovarian cancer, primary peritoneal carcinoma or fallopian tube cancer of high grade serous and high grade endometrioid subtypes. Patients who have a diagnosis of ovarian cancer with a known germline mutation in *BRCA1* or *BRCA2* will also be eligible for inclusion regardless of histological subtype. Patients who are having a diagnostic image-guided biopsy may be consented and study biopsy taken while awaiting pathological review. Eligible patients who have had samples collected under generic research consent may be registered retrospectively only after full discussion with the site, Chief Investigator and CR-UK CTU (and BriTROC-1 specific consent obtained)	8
Has the patient received at least one line of platinum-containing chemotherapy	0
Please confirm that there is formalin-fixed, paraffin embedded tissue taken at the time of original diagnosis of high grade serous ovarian cancer available?	11
Does the patient have disease deemed suitable for imaging-guided/intra-operative or other suitable biopsy	90
Is the patient able to comply with study procedures?	11
Does the patient have a life expectancy of > 3 months?	0
Is it confirmed that there is no contraindication to biopsy?	23
Does the patient have an ovarian, primary peritoneal or fallopian tube cancer of low grade serous, grades 1 or 2 endometrioid, clear cell or carcinosarcoma/MMMT subtypes unless associated with known germline mutation *BRCA1* or *BRCA2*	10
Does the patient have a borderline/low malignant potential tumour?	0
Does the patient have a non-epithelial ovarian malignancy	0
Does the patient have an asymptomatic rising CA125 with no radiological evidence of recurrent ovarian cancer?	4
Was the original diagnosis of high grade serous cancer made on cytology only?	9

Not stated	16
TOTAL	**182**
Declined – other reason	
Patient did not want biopsy	5
Commencing chemotherapy treatment asap	12
Felt she had 'too much going on'	8
Patient opted for treatment with another study	1
Patient prefers supportive care locally therefore not approached	1
Did not wish to undergo biopsy due to previous painful biopsy experience at another hospital	2
Patient did not want extra tests/visits	4
Patient too unwell for chemo or biopsy	2
Patients family not happy for her to participate	2
Patient accepted but started treatment sooner than planned	1
Area to be biopsied too painful	1
Patient cannot commit	1
Patient too distressed by progression	2
Not well enough	2
Patient's husband is ill	2
Psychologically not able to cope with anything else (depression)	1
Not happy to take part in any research	1
Patient not feeling up to it	2
Unable to attend for biopsy on only day available	1
Not interested in participation	2
Biopsy at diagnosis and found it very traumatic	1
Eligible but having urgent radiotherapy	1
Patient discussed at MDT - Research Nurses not informed - pt proceeded to surgery	2
Consultant decided trial not appropriate	1
Patient transferred to another hospital	3

Patient entered another trial	2
Unspecified	11
No information received	5
TOTAL	**79**

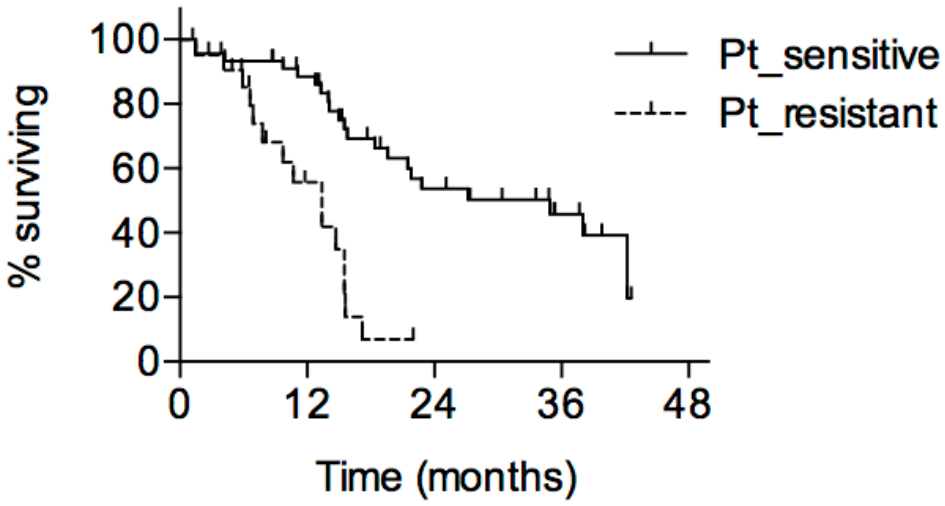
4. Prior chemotherapy lines and time from diagnosis to recruitment

Time to recruitment from diagnosis for platinum sensitive relapse (left) and platinum resistant relapse (right) patients, grouped by lines of prior chemotherapy. Each point (•) represents one patient.



5. Overall survival

Overall survival of patients recruited in Glasgow (n=52) and Edinburgh (n=16) from time of consent. Pt_sensitive - platinum-sensitive (relapse ≥6 months following previous platinum-based chemotherapy). Pt_resistant (relapse <6 months following previous platinum-based chemotherapy.



6. Adverse events

Trial No	Complications Following	Biopsy type	AE	Grade	Relationship To Biopsy
1	Baseline biopsy	Image-guided biopsy	Pain	2	Probable
14	Baseline biopsy	Image-guided biopsy	Pain	1	Probable
14	2nd biopsy	Image-guided biopsy	Postoperative Haemorrhage (post biopsy)	1	Probable
14	2nd biopsy	Image-guided biopsy	Pain	1	Probable
16	Baseline biopsy	Image-guided biopsy	Pain	2	Possible
25	Baseline biopsy	Image-guided biopsy	Postoperative Haemorrhage (post biopsy)	2	Definite
25	Baseline biopsy	Image-guided biopsy	Pain	2	Definite
54	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
59	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
79	Baseline biopsy	Image-guided biopsy	General disorders and administration site conditions – Other (haematoma)	1	Definite
106	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
116	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
123	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
130	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
136	Baseline biopsy	Image-guided biopsy	Pain	1	Definite
148	Baseline biopsy	Image-guided biopsy	Pain	1	Probable
179	Baseline biopsy	Image-guided biopsy	Postoperative Haemorrhage (post biopsy)	1	Definite
179	Baseline biopsy	Image-guided biopsy	Vaginal discharge	2	Definite
183	Baseline biopsy	Image-guided	Pain	1	Definite

		biopsy			
192	Baseline biopsy	Image-guided biopsy	Pain	1	Possible
201	Baseline biopsy	Image-guided biopsy	Postoperative Haemorrhage (post biopsy)	1	Definite
211	Baseline biopsy	Image-guided biopsy	Pain	1	Definite

7. Patient consent form

CONSENT FORM FOR PATIENTS/VOLUNTEERS IN A CLINICAL RESEARCH PROJECT
(Form to be on hospital headed paper)

Title of Project: BriTROC-1: sample collection study to investigate the role of Homologous Recombination Deficiency in platinum sensitivity in recurrent high grade serous ovarian cancer

Principal Investigator:
Patient Study Number:

		Please initial box
1.	I confirm that I have read and understand the information sheet dated 26th April 2016 (Version 9) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected	
3.	I agree that relevant sections of any of my medical notes and data collected during the study may be looked at by responsible individuals from the Cancer Research UK Clinical Trials Unit (Glasgow), the trial sponsor, the regulatory authorities and the NHS organisation where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.	
4.	I agree to give tissue samples from my diagnosis for use in research related to this study as described in the patient information sheet.	
5.	I agree to give a 40 ml blood samples for research purposes as described in the patient information sheet.	
6.	I agree to give an ascitic fluid sample for research purposes if ascites is present and drainage is clinically indicated, as described in the patient information sheet.	
7.	I agree to the information detailed in this patient information sheet to be collected as part of this study.	
8.	I understand that the aforementioned samples will be held at the Institute for Cancer Sciences, University of Glasgow.	
9.	I understand that my samples may be retained by the Institute for Cancer Sciences, University of Glasgow, for use in future biological studies to help understand ovarian cancer. I understand that some of these projects may be carried out by researchers other than the BriTROC co-investigators. I understand that some of these projects may involve international collaborations. All future research will be approved by a research ethics committee.	
10.	I agree to my GP being informed of my participation in this trial.	
11.	I agree that my routine data that are collected nationally, as described in the patient information sheet, can be accessed.	
12.	I agree to take part in this study	

	OPTIONAL	**Please initial box**	
13.	I agree to two extra blood samples, taken just before first and second cycle of chemotherapy/anti-cancer treatment	**Yes**	**No**
14.	If my cancer returns during the duration of this study, I agree to undergo a repeat biopsy and to give 40 ml blood samples for research purposes as described in the patient information sheet	**Yes**	**No**
15.	I wish to be informed of any inherited mutations in DNA repair genes found during this research project and consent to a blood sample being sent for storage at the regional NHS Genetics Laboratory for confirmatory testing if appropriate. If you or your family wish to be referred to the Clinical Genetics Service this can be arranged through your doctor.	**Yes**	**No**
16.	In the event that I am no longer alive, I wish my next-of-kin to be informed of any inherited mutations in DNA repair genes found during this research project	**Yes**	**No**
17.	I give permission for my initials, date of birth and NHS (or Community Health Index (CHI)) number to be collected by the Cancer Research UK Clinical Trials Unit (Glasgow), where they will be stored in a secure location. I understand giving consent to the use of this data as described is optional and not mandatory for participating in this study.	**Yes**	**No**

Name of Patient	Date	Signature
_____	_____	_____
Name of person taking consent (designated responsible person)	Date	Signature
_____	_____	_____

When completed, 1 original for patient, 1 original for researcher, 1 copy to be kept with hospital notes

8. R Markdown

Teodora Goranova, Iain A. McNeish, James D. Brenton

2017-02-27

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Please use the Google R style guide, see https://google.github.io/styleguide/Rguide.xml.

We first load required libraries and define Useful functions.

Load data files

We first load the raw data from 6 tables and derive factors.

```
## Parsed with column specification:
## cols(
##   trialno = col_double(),
##   platinum.status = col_character(),
##   site = col_character(),
##   reg.date = col_datetime(format = ""),
##   diag.age = col_double(),
##   study.age = col_double(),
##   histo = col_character(),
##   diagdate = col_datetime(format = ""),
##   time.since.diag = col_double(),
##   prior.lines.chemo = col_integer(),
##   registration = col_character(),
##   brca.status = col_character(),
##   censor.date = col_datetime(format = ""),
##   os.status = col_integer(),
##   os.diag.months = col_double(),
##   os.entry.months = col_double()
## )

## Parsed with column specification:
## cols(
##   trialno = col_double(),
##   start.date = col_datetime(format = ""),
##   end.date = col_datetime(format = ""),
##   numberof.cycles = col_double(),
##   line.no = col_double(),
##   drug = col_character(),
##   dose = col_double(),
##   units = col_character(),
##   comment = col_character()
## )
```

```
## Parsed with column specification:
## cols(
##   trialno = col_double(),
##   biopsy.sequence = col_character(),
##   ae = col_character(),
##   grade = col_double(),
##   relationship.to.biopsy = col_character(),
##   ae.details = col_character()
## )

## Parsed with column specification:
## cols(
##   trialno = col_double(),
##   biopsy.type = col_character(),
##   needle.size = col_character(),
##   biopsy.site = col_character(),
##   biopsy.sequence = col_character(),
##   needle.size.bin = col_character(),
##   biopsy.site.recode = col_character()
## )

## Parsed with column specification:
## cols(
##   trialno = col_double(),
##   histo.no = col_character(),
##   jblab.id = col_character(),
##   biorepository.no = col_double(),
##   biorepository.no.h.e = col_character(),
##   nbf.umfix = col_character(),
##   tumour.cellularity.comments = col_character(),
##   tumour.cellularity = col_double(),
##   tumour.cellularity.in.dissected.area = col_double(),
##   dissection.method = col_character(),
##   sections.for.dna.extraction = col_double(),
##   dna.extraction.date = col_date(format = ""),
##   kit.used.for.extraction = col_character(),
##   elution.buffer = col_character(),
##   dna.qubit.ng.ul = col_double(),
##   volume.ul = col_double(),
##   total.dna.ug = col_double(),
##   comments.1 = col_character(),
##   biopsy.sequence = col_character()
## )

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   trialno = col_integer(),
##   chromosome = col_integer(),
##   position = col_integer(),
##   af1 = col_double(),
##   af2 = col_double(),
##   depth1 = col_integer(),
```

```
##   depth2 = col_integer(),
##   quality.1 = col_integer(),
##   quality.2 = col_integer(),
##   offset.from.primer.end = col_integer(),
##   indel.length = col_integer(),
##   af1.short = col_double(),
##   af2.short = col_double(),
##   depth1.short = col_integer(),
##   depth2.short = col_integer(),
##   mean.af = col_double(),
##   mean.depth = col_double()
## )

## See spec(...) for full column specifications.
```

Patients

Table 1: Demographic and disease characteristics

Total number of patients

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   220
```

Are there any duplicate trial no?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   220

## # A tibble: 2 × 3
##    registration     n rel.freq
##           <chr> <int>    <dbl>
## 1   Prospective   198      0.9
## 2 Retrospective    22      0.1
```

How many NA are there in the clin.data table?

```
## # A tibble: 1 × 16
##   trialno platinum.status  site reg.date diag.age study.age histo diagdate
##     <int>           <int> <int>    <int>    <int>     <int> <int>    <int>
## 1       0               0     0        0        0         2     0        2
## # ... with 8 more variables: time.since.diag <int>,
## #   prior.lines.chemo <int>, registration <int>, brca.status <int>,
## #   censor.date <int>, os.status <int>, os.diag.months <int>,
## #   os.entry.months <int>
```

How many platinum sensitive vs. resistant?

```
## # A tibble: 2 × 3
##                         platinum.status     n  rel.freq
##                                   <chr> <int>     <dbl>
## 1 Platinum resistant/refractory relapse    49 0.2227273
## 2            Platinum sensitive relapse   171 0.7772727
```

Key variables by platinum sensitivity.

```
## # A tibble: 2 × 10
##                         platinum.status study.age_median_ study.age_min_
##                                   <chr>             <dbl>          <dbl>
## 1 Platinum resistant/refractory relapse          65.84658       25.04110
## 2            Platinum sensitive relapse          69.44658       37.18904
## # ... with 7 more variables: study.age_max_ <dbl>,
## #   time.since.diag_median_ <dbl>, time.since.diag_min_ <dbl>,
## #   time.since.diag_max_ <dbl>, prior.lines.chemo_median_ <int>,
## #   prior.lines.chemo_min_ <int>, prior.lines.chemo_max_ <int>
```

Key variables in whole group.

```
## # A tibble: 1 × 9
##   study.age_median_ study.age_min_ study.age_max_ time.since.diag_median_
##               <dbl>          <dbl>          <dbl>                   <dbl>
## 1          68.21644        25.0411       92.52329                31.00275
## # ... with 5 more variables: time.since.diag_min_ <dbl>,
## #   time.since.diag_max_ <dbl>, prior.lines.chemo_median_ <dbl>,
## #   prior.lines.chemo_min_ <int>, prior.lines.chemo_max_ <int>
```

What was BRCA status by platinum status?

```
## Source: local data frame [10 x 4]
## Groups: platinum.status [2]
##
##                          platinum.status brca.status     n  rel.freq
##                                    <chr>       <chr> <int>     <dbl>
## 1             Platinum sensitive relapse       BRCA1    11  6.432749
## 2             Platinum sensitive relapse       BRCA2     7  4.093567
## 3             Platinum sensitive relapse     Missing    63 36.842105
## 4             Platinum sensitive relapse  Not tested    71 41.520468
## 5             Platinum sensitive relapse   Wild-type    19 11.111111
## 6  Platinum resistant/refractory relapse       BRCA1     3  6.122449
## 7  Platinum resistant/refractory relapse       BRCA2     5 10.204082
## 8  Platinum resistant/refractory relapse     Missing    14 28.571429
## 9  Platinum resistant/refractory relapse  Not tested    13 26.530612
## 10 Platinum resistant/refractory relapse   Wild-type    14 28.571429
```

What was BRCA status across whole group?

```
## # A tibble: 5 × 3
##   brca.status     n  rel.freq
##         <chr> <int>     <dbl>
## 1       BRCA1    14  6.363636
## 2       BRCA2    12  5.454545
## 3     Missing    77 35.000000
```

```
## 4  Not tested    84 38.181818
## 5   Wild-type    33 15.000000
```

What were histologies by platinum status?

```
## Source: local data frame [5 x 4]
## Groups: platinum.status [2]
##
##                         platinum.status                  histo     n
##                                   <chr>                  <chr> <int>
## 1            Platinum sensitive relapse      High grade serous   160
## 2            Platinum sensitive relapse Endometrioid (grade 3)     5
## 3            Platinum sensitive relapse                Missing     5
## 4            Platinum sensitive relapse         Carcinosarcoma     1
## 5 Platinum resistant/refractory relapse      High grade serous    49
## # ... with 1 more variables: rel.freq <dbl>
```

What were histologies across whole group?

```
## # A tibble: 4 × 3
##                    histo     n   rel.freq
##                    <chr> <int>      <dbl>
## 1      High grade serous   209 95.0000000
## 2 Endometrioid (grade 3)     5  2.2727273
## 3                Missing     5  2.2727273
## 4         Carcinosarcoma     1  0.4545455
```

What were prior treatment regimes by platinum status?

```
## Source: local data frame [13 x 4]
## Groups: platinum.status [2]
##
##                          platinum.status prior.lines.chemo     n
##                                    <chr>             <int> <int>
## 1             Platinum sensitive relapse                 1   120
## 2             Platinum sensitive relapse                 2    39
## 3             Platinum sensitive relapse                 3     6
## 4             Platinum sensitive relapse                 4     1
## 5             Platinum sensitive relapse                 5     3
## 6             Platinum sensitive relapse                NA     2
## 7  Platinum resistant/refractory relapse                 1    17
## 8  Platinum resistant/refractory relapse                 2    23
## 9  Platinum resistant/refractory relapse                 3     1
## 10 Platinum resistant/refractory relapse                 4     2
## 11 Platinum resistant/refractory relapse                 5     4
## 12 Platinum resistant/refractory relapse                 6     1
## 13 Platinum resistant/refractory relapse                12     1
## # ... with 1 more variables: rel.freq <dbl>
```

What were prior treatment regimes across whole group?

```
## # A tibble: 8 × 3
##   prior.lines.chemo     n   rel.freq
##               <int> <int>      <dbl>
```

```
## 1                 1   137 62.2727273
## 2                 2    62 28.1818182
## 3                 3     7  3.1818182
## 4                 4     3  1.3636364
## 5                 5     7  3.1818182
## 6                 6     1  0.4545455
## 7                12     1  0.4545455
## 8                NA     2  0.9090909
```

Biopsy features

How many biopsies were performed?

```
## Source: local data frame [7 x 3]
## Groups: biopsy.type [?]
##
##    biopsy.type biopsy.sequence     n
##          <chr>           <chr> <int>
## 1 image-guided        baseline   118
## 2 image-guided          second     7
## 3      missing        baseline     1
## 4        other        baseline     1
## 5        other          second     1
## 6     surgical        baseline    96
## 7     surgical          second     3
```

Did any patient have more than one sample obtained at baseline and second image-guided biopsy?

```
## Source: local data frame [125 x 3]
## Groups: trialno [119]
##
##    trialno biopsy.sequence     n
##      <dbl>           <chr> <int>
## 1        1        baseline     1
## 2        2        baseline     1
## 3        4        baseline     1
## 4        5        baseline     1
## 5        6        baseline     1
## 6        7        baseline     1
## 7        8        baseline     1
## 8        9        baseline     1
## 9       10        baseline     1
## 10      12        baseline     1
## # ... with 115 more rows
```

What needle size was used for image-guided biopsies (baseline and second)?

```
## # A tibble: 3 × 3
##   needle.size     n     freq
##         <chr> <int>    <dbl>
## 1         18G    54 45.00000
```

```
## 2         16G    49 40.83333
## 3         14G    17 14.16667
```

Table 2: Biopsy locations

What were the biopsy sites?

```
## Source: local data frame [49 x 5]
## Groups: biopsy.type, biopsy.sequence [7]
##
##     biopsy.type biopsy.sequence           biopsy.site     n      freq
##           <chr>           <chr>                 <chr> <int>     <dbl>
## 1  image-guided        baseline       peritoneum.abdo    27 22.881356
## 2  image-guided        baseline             ln.other    26 22.033898
## 3  image-guided        baseline                liver    13 11.016949
## 4  image-guided        baseline              omentum    12 10.169492
## 5  image-guided        baseline subcut.abdochest.wall     9  7.627119
## 6  image-guided        baseline     peritoneum.pelvic     7  5.932203
## 7  image-guided        baseline             ln.retrop     6  5.084746
## 8  image-guided        baseline             ln.pelvis     5  4.237288
## 9  image-guided        baseline            vag.vault     4  3.389831
## 10 image-guided        baseline        uterus.ov.tube     2  1.694915
## # ... with 39 more rows
```

What was the frequency by binned biopsy sites?

```
## # A tibble: 10 × 3
##       biopsy.site.recode     n      freq
##                    <chr> <int>     <dbl>
## 1             lymph.node    64 28.193833
## 2             peritoneum    53 23.348018
## 3                omentum    26 11.453744
## 4                  other    19  8.370044
## 5                  liver    17  7.488987
## 6            gynae.organ    14  6.167401
## 7  subcut.abdochest.wall    14  6.167401
## 8    bowel.serosa.mesent    12  5.286344
## 9              diaphragm     4  1.762115
## 10          peri.splenic     4  1.762115
```

Which biopsy sites are in the NA bin?

```
## # A tibble: 0 × 2
## # ... with 2 variables: biopsy.site <chr>, n <int>
```

Which sites are included in gynae.organ bin?

```
## # A tibble: 2 × 2
##      biopsy.site     n
##            <chr> <int>
## 1 uterus.ov.tube     6
## 2      vag.vault     8
```

Table 3: Adverse events

We use a left join to join the adverse event table to the biopsies table, so all adverse events have biopsy details.

```
## Joining, by = c("trialno", "biopsy.sequence")
```

How many biopsies have adverse events?

```
## Source: local data frame [5 x 3]
## Groups: biopsy.type [?]
##
##    biopsy.type biopsy.sequence     n
##          <chr>           <chr> <int>
## 1 image-guided        baseline    20
## 2 image-guided          second     2
## 3        other        baseline     1
## 4     surgical        baseline    14
## 5         <NA>        baseline     1
```

How many patients had adverse events after both baseline and second image-guided biopsy?

```
## Source: local data frame [5 x 3]
## Groups: biopsy.type [?]
##
##    biopsy.type biopsy.sequence no.patients
##          <chr>           <chr>       <int>
## 1 image-guided        baseline          18
## 2 image-guided          second           1
## 3        other        baseline           1
## 4     surgical        baseline           9
## 5         <NA>        baseline           1
```

What adverse events occured after image-guided biopsy?

```
## Source: local data frame [6 x 3]
## Groups: ae [?]
##
##                                                                      ae
##                                                                   <chr>
## 1 General disorders and administration site conditions - Other, specify
## 2                                                                  Pain
## 3                                                                  Pain
## 4                              Postoperative Haemorrhage (post biopsy)
## 5                              Postoperative Haemorrhage (post biopsy)
## 6                                                     Vaginal discharge
## # ... with 2 more variables: grade <dbl>, n <int>
```

What adverse events occurred by needle size?

```
## Source: local data frame [10 x 4]
## Groups: ae, needle.size [?]
##
```

```
##                                                                       ae
##                                                                    <chr>
## 1  General disorders and administration site conditions - Other, specify
## 2                                                                    Pain
## 3                                                                    Pain
## 4                                                                    Pain
## 5                                                                    Pain
## 6                                                                    Pain
## 7                                 Postoperative Haemorrhage (post biopsy)
## 8                                 Postoperative Haemorrhage (post biopsy)
## 9                                 Postoperative Haemorrhage (post biopsy)
## 10                                                        Vaginal discharge
## # ... with 3 more variables: needle.size <chr>, grade <dbl>, n <int>
```

Were adverse events different between needle size?

```
## Joining, by = c("trialno", "biopsy.sequence")
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  foo$needle.size.bin and foo$ae.bin
## X-squared = 0.22802, df = 1, p-value = 0.633
```

Quality of biopsies

DNA extraction

We left join the tissue.dna data to the biopsies data.

```
## Joining, by = c("trialno", "biopsy.sequence")
```

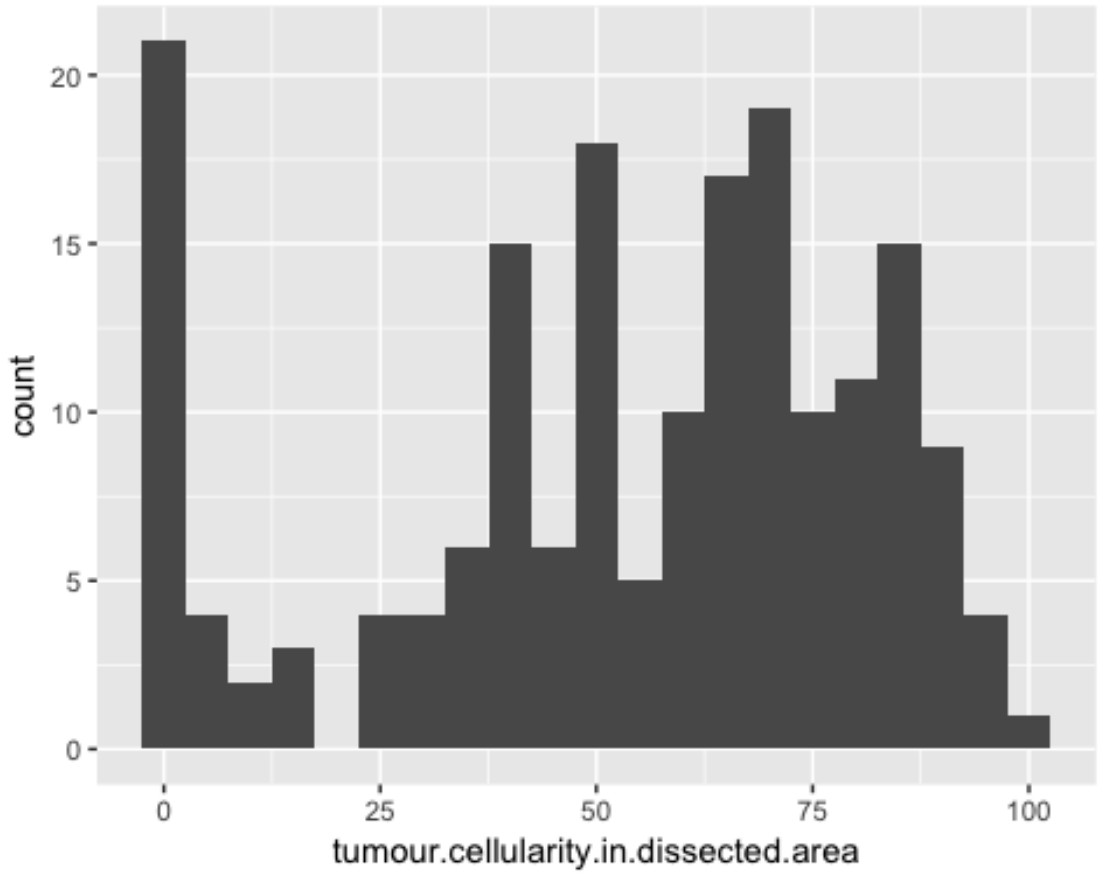
How many unique patients?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   142
```

How many unique samples had DNA extraction attempted?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   184
```

What was the cellularity of dissected tumour area (irrespective of yield)?



How many dissected samples had cellularity > 0?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   163
```

What was dissection method for all samples?

```
## # A tibble: 3 × 2
##   dissection.method     n
##               <chr> <int>
## 1             Macro    41
## 2             Micro   122
## 3              <NA>    21
```

NA here stand for samples with no tumour cells.

What was the type of samples that had cellularity > 0?

```
## # A tibble: 2 × 3
##    biopsy.type     n     freq
##          <chr> <int>    <dbl>
## 1 image-guided    88 53.98773
## 2     surgical    75 46.01227
```

What was the yield and cellularity in samples with cellularity > 0?

```
## # A tibble: 1 × 10
##   total.dna.ug_min_ cellularity_min_ total.dna.ug_Q1_ cellularity_Q1_
##               <dbl>            <dbl>            <dbl>           <dbl>
## 1                 0                5            0.844              45
## # ... with 6 more variables: total.dna.ug_median_ <dbl>,
## #   cellularity_median_ <dbl>, total.dna.ug_Q3_ <dbl>,
## #   cellularity_Q3_ <dbl>, total.dna.ug_max_ <dbl>, cellularity_max_ <dbl>
```

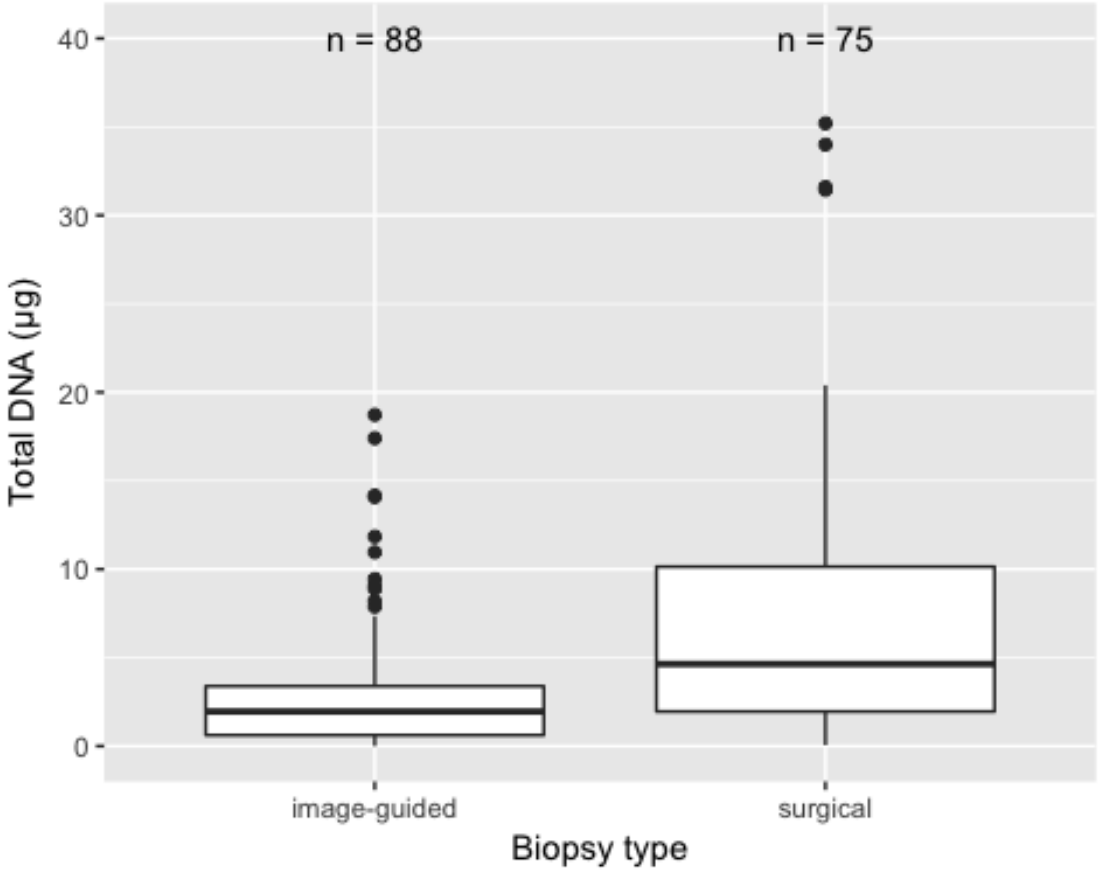
What was the yield by biopsy type in tumour samples?

```
## # A tibble: 2 × 6
##    biopsy.type   min_   Q1_ median_    Q3_  max_
##          <chr>  <dbl> <dbl>   <dbl>  <dbl> <dbl>
## 1 image-guided 0.0000 0.621   1.955  3.385 18.72
## 2     surgical 0.0556 1.956   4.640 10.140 35.20
```

Plot DNA yield by biopsy type

```
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).

## Warning: Removed 21 rows containing non-finite values (stat_summary).
```



Does the DNA yield differ by the type of biopsy?

```
##   statistic     p.value                                                 method
## 1    1919.5 4.32916e-06 Wilcoxon rank sum test with continuity correction
```

```
##   alternative
## 1   two.sided
```

How many samples with > 200ng DNA?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   158
```

How many patients with a sample > 200ng?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   124
```

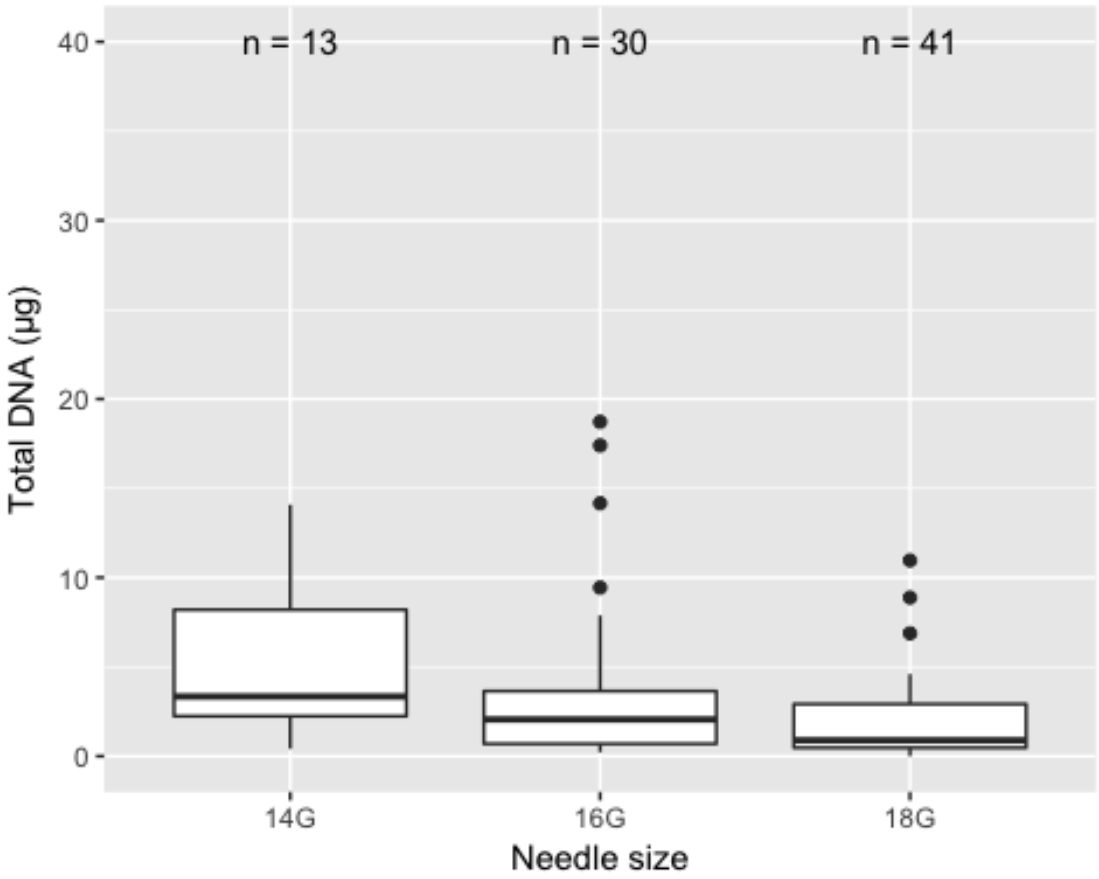
What was dissection method for samples with > 200ng of DNA?

```
## # A tibble: 2 × 2
##   dissection.method     n
##               <chr> <int>
## 1             Macro    40
## 2             Micro   118
```

What is the yield by needle size?

```
## Warning: Removed 14 rows containing non-finite values (stat_boxplot).

## Warning: Removed 14 rows containing non-finite values (stat_summary).
```



We summarize yield using the derived column "needle.size.bin"

```
## # A tibble: 2 × 6
##   needle.size.bin  min_   Q1_ median_   Q3_  max_
##             <chr> <dbl> <dbl>   <dbl> <dbl> <dbl>
## 1         14G+16G 0.222 0.848   2.860  5.54 18.72
## 2             18G 0.000 0.472   0.888  2.92 10.96
```

Are differences in yield from needle size significant? Test hypothesis with non-parametric test.

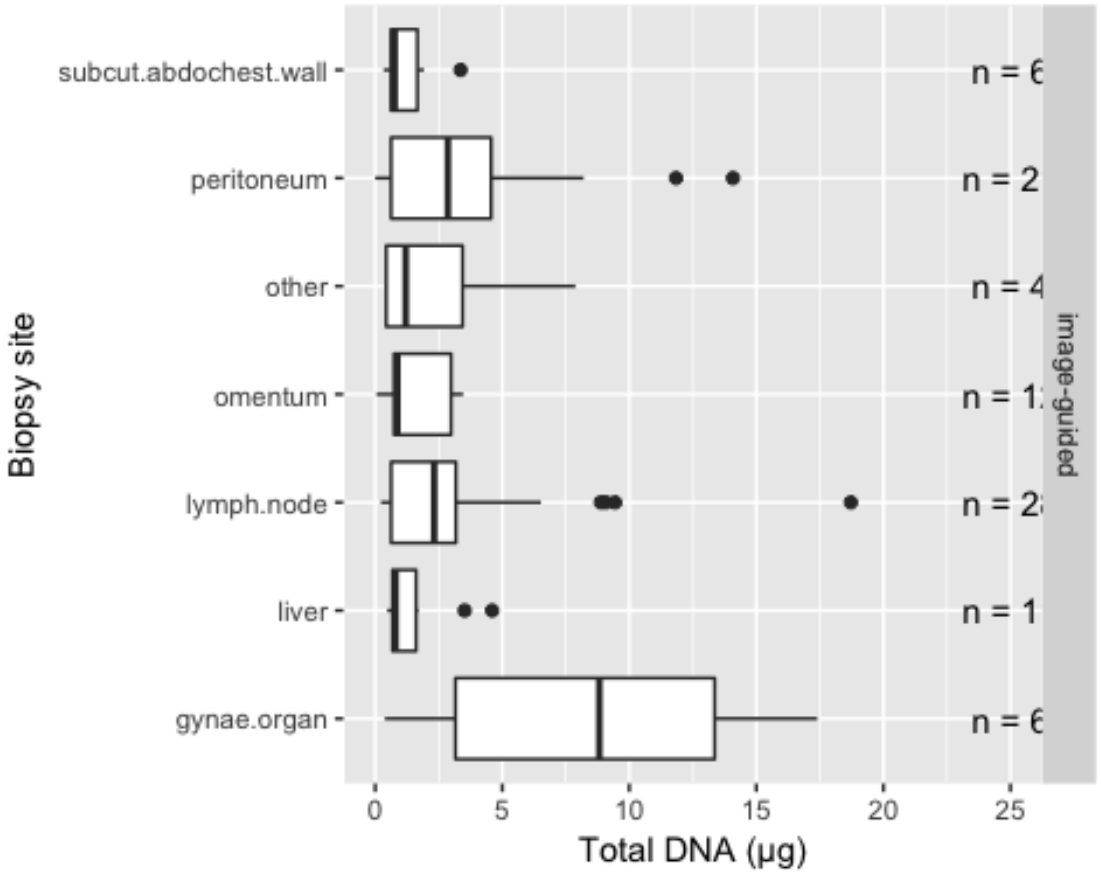
```
##   statistic    p.value                                                 method
## 1    1167.5 0.01062334 Wilcoxon rank sum test with continuity correction
##   alternative
## 1   two.sided
```

What was the yield by biopsy location using recoded sites (both surgical and image guided)?

```
## Source: local data frame [16 x 7]
## Groups: biopsy.type [?]
##
##     biopsy.type    biopsy.site.recode   min_    Q1_ median_     Q3_  max_
##           <chr>                 <chr>  <dbl>  <dbl>   <dbl>   <dbl> <dbl>
## 1  image-guided           gynae.organ 0.3800 3.1580  8.8200 13.3600 17.40
## 2  image-guided                 liver 0.4720 0.6815  0.8000  1.6080  4.60
## 3  image-guided            lymph.node 0.2220 0.6270  2.3200  3.1750 18.72
## 4  image-guided               omentum 0.0520 0.7410  0.8700  2.9850  3.46
```

```
## 5  image-guided                other 0.4040 0.4340  1.2000  3.4370  7.88
## 6  image-guided           peritoneum 0.0000 0.6240  2.8600  4.5600 14.08
## 7  image-guided subcut.abdochest.wall 0.3300 0.6200  0.7660  1.6590  3.36
## 8      surgical   bowel.serosa.mesent 0.0800 1.8320  6.3200  7.8800 11.76
## 9      surgical             diaphragm 0.0556 0.8117  1.5678  2.3239  3.08
## 10     surgical          gynae.organ 0.4400 0.6480  2.1000 13.3200 19.44
## 11     surgical           lymph.node 0.5160 1.6150  3.8700 15.0600 35.20
## 12     surgical              omentum 0.9320 1.4650  4.5560  7.6800 14.64
## 13     surgical                other 1.0800 4.6400  6.2400  7.9900 18.16
## 14     surgical         peri.splenic 2.6000 3.6800  5.2000  8.0900 13.28
## 15     surgical           peritoneum 3.3000 5.7600 10.0600 11.9300 20.40
## 16     surgical subcut.abdochest.wall 0.5120 3.9600  4.8000 12.2000 20.40
```

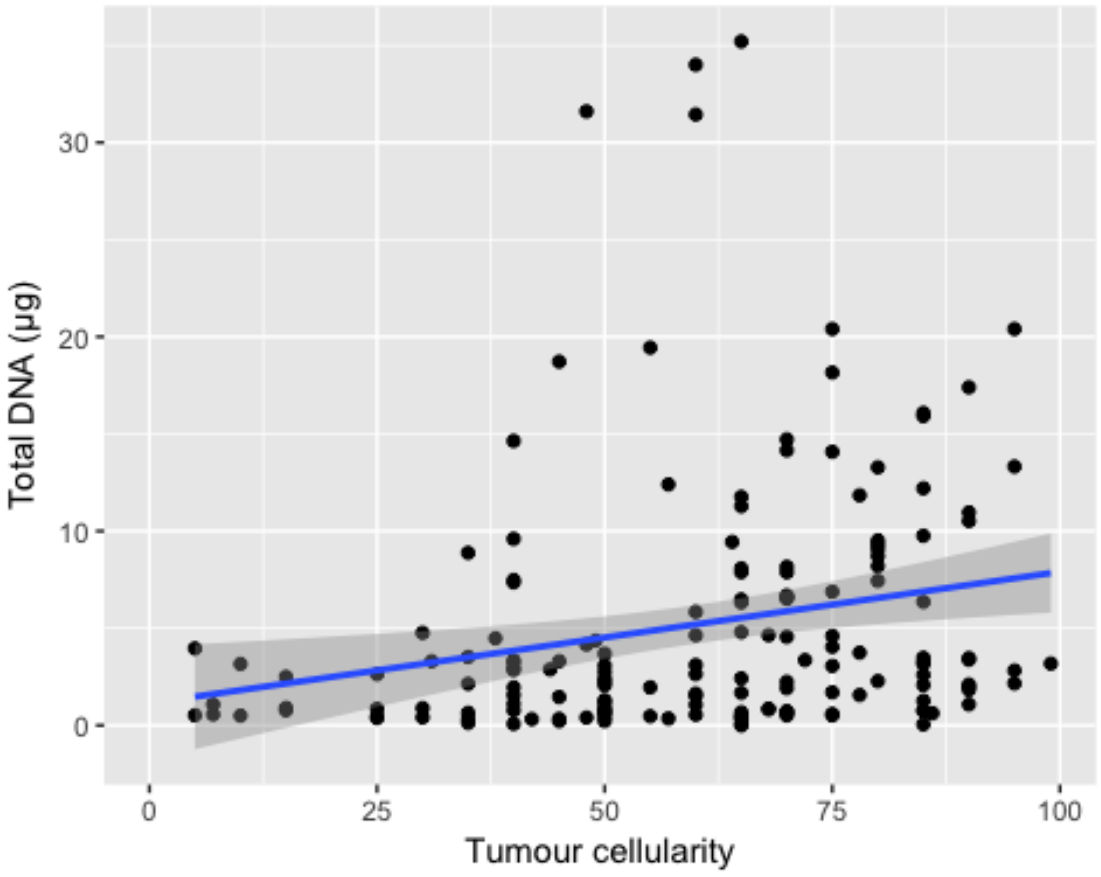
Plot data for total DNA by tissue site.



Are cellularity and total DNA yield correlated?

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).

## Warning: Removed 21 rows containing missing values (geom_point).
```



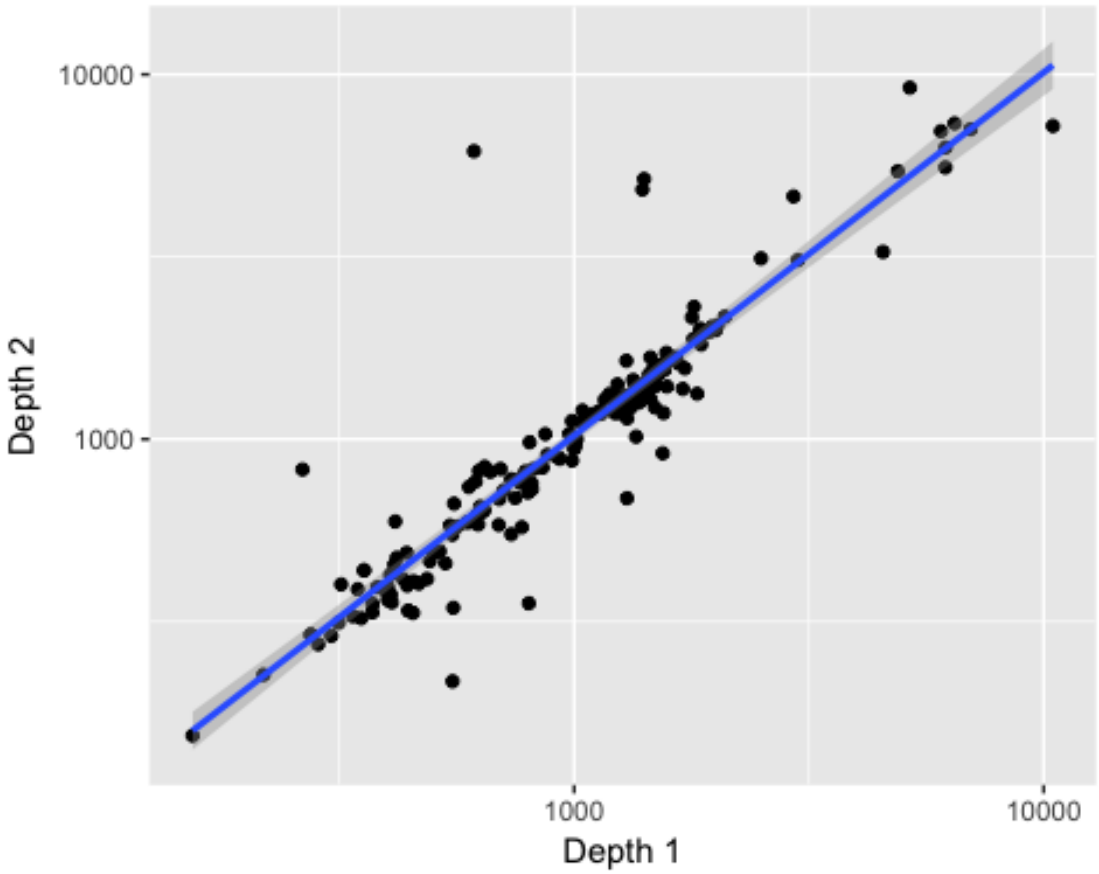

```
##    estimate statistic      p.value                          method
## 1 0.3044246  502041.3 7.785379e-05 Spearman's rank correlation rho
##   alternative
## 1   two.sided
```

TAm-seq

Are the technical replicates consistent? For sequencing depth?

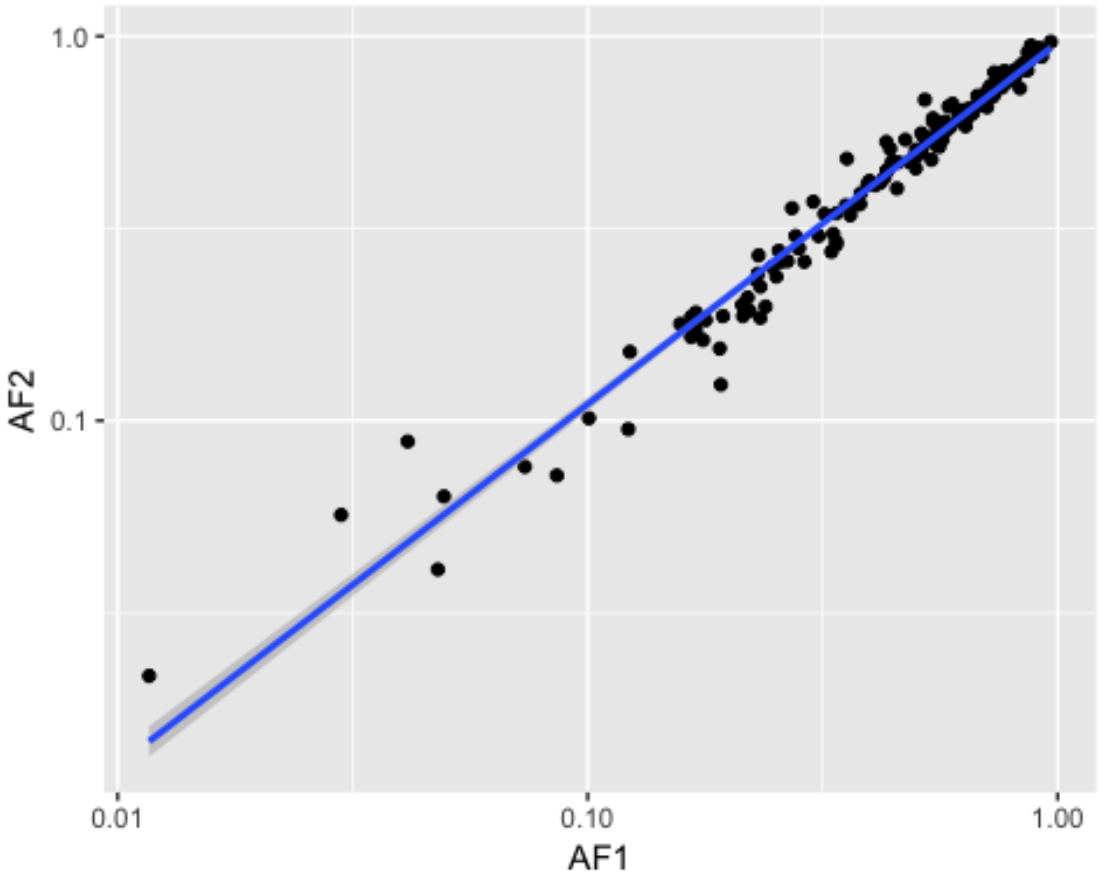
```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).

## Warning: Removed 7 rows containing missing values (geom_point).
```



For allele fraction?

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
## Warning: Removed 7 rows containing missing values (geom_point).
```



What was the sequencing depth and allele fraction (using mean of technical replicates)?

```
## # A tibble: 1 × 6
##   mean.depth_median_ mean.af_median_ mean.depth_min_ mean.af_min_
##                <dbl>           <dbl>           <dbl>        <dbl>
## 1            1086.25         0.55375             154   0.01666992
## # ... with 2 more variables: mean.depth_max_ <dbl>, mean.af_max_ <dbl>
```

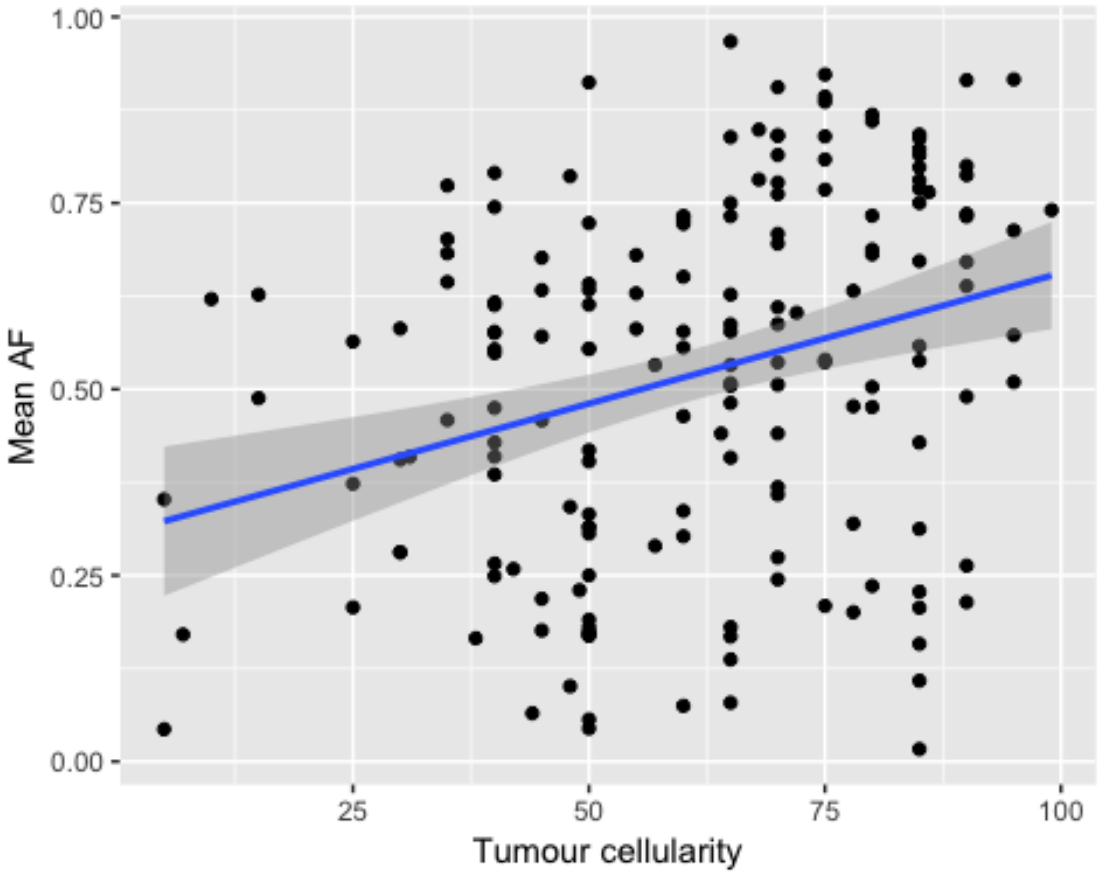
We join tam.seq data to the dna and biopsy data.

```
## Joining, by = c("jblab.id", "trialno")
```

Are cellularity and mean allele fraction related?

```
## Warning: Removed 8 rows containing non-finite values (stat_smooth).

## Warning: Removed 8 rows containing missing values (geom_point).
```

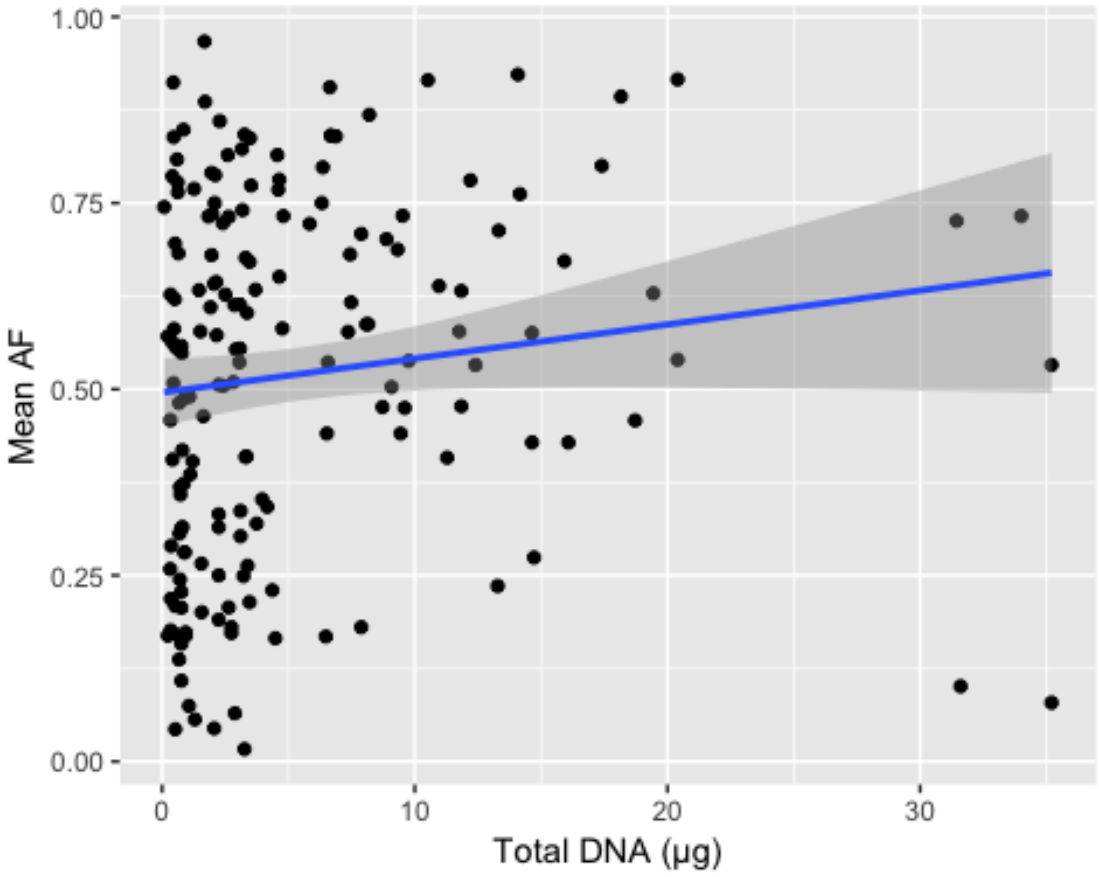



```
##    estimate statistic      p.value                          method
## 1 0.3142648    623059 2.160671e-05 Spearman's rank correlation rho
##   alternative
## 1   two.sided
```

Are DNA yield and mean allele fraction related?

```
## Warning: Removed 8 rows containing non-finite values (stat_smooth).

## Warning: Removed 8 rows containing missing values (geom_point).
```




```
##    estimate statistic     p.value                          method
## 1 0.2199494    708754 0.003355179 Spearman's rank correlation rho
##   alternative
## 1   two.sided
```

How many unique patients had tagged amplicon sequencing?

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   125
```

How many patients had a mutation?

```
## # A tibble: 5 × 2
##     gene     n
##    <chr> <int>
## 1   KRAS     6
## 2 PIK3CA     3
## 3   PTEN     2
## 4   TP53   118
## 5   <NA>     4
```

How many unique samples had tagged amplicon sequencing?

```
## # A tibble: 1 × 1
##       n
```

```
##   <int>
## 1   160
```

Which cases had multiple samples with TP53 mutations?

```
## # A tibble: 118 × 2
##    trialno     n
##      <int> <int>
## 1       32     7
## 2       27     4
## 3       45     4
## 4       74     4
## 5       88     4
## 6       94     4
## 7      103     4
## 8        1     3
## 9        2     2
## 10       5     2
## # ... with 108 more rows
```

What is proportion of binned TP53 mutations?

```
## # A tibble: 3 × 3
##   tp53.type.recode     n      freq
##              <chr> <int>     <dbl>
## 1    nonsynonymous    71 60.169492
## 2       truncating    41 34.745763
## 3          inframe     6  5.084746
```

What is proportion of all TP53 mutation types?

```
## # A tibble: 9 × 3
##                   type     n       freq
##                  <chr> <int>      <dbl>
## 1        nonsynonymous    71 60.1694915
## 2       frameshift_del    14 11.8644068
## 3      splice_acceptor     8  6.7796610
## 4       frameshift_ins     7  5.9322034
## 5          stop_gained     7  5.9322034
## 6     inframe_deletion     5  4.2372881
## 7         splice_donor     4  3.3898305
## 8    inframe_insertion     1  0.8474576
## 9 splice_region,intron     1  0.8474576
```