

# Multisite analysis of high-grade serous epithelial ovarian cancers identifies genomic regions of focal and recurrent copy number variation in 3p26.2 and 8q24.3

Sara Ballabio<sup>1\*</sup>, Iliara Craparotta<sup>1\*</sup>, Lara Paracchini<sup>1\*</sup>, Laura Mannarino<sup>1</sup>, Silvia Corso<sup>2</sup>, Maria Grazia Pezzotta<sup>3</sup>, Martina Vescio<sup>1,7</sup>, Robert Fruscio<sup>4</sup>, Chiara Romualdi<sup>5</sup>, Emanuele Dainese<sup>3</sup>, Lorenzo Ceppi<sup>4</sup>, Enrica Calura<sup>5</sup>, Silvana Pileggi<sup>1</sup>, Giulia Siravegna<sup>6</sup>, Linda Pattini<sup>7</sup>, Paolo Martini<sup>5</sup>, Martina delle Marchette<sup>4</sup>, Costantino Mangioni<sup>2</sup>, Antonio Ardizzoia<sup>3</sup>, Antonio Pellegrino<sup>2</sup>, Fabio Landoni<sup>4</sup>, Maurizio D’Incalci<sup>1#</sup>, Luca Beltrame<sup>1§</sup>, Sergio Marchini<sup>1§</sup>

## Authors’ Affiliation

<sup>1</sup>Department of Oncology, IRCCS Istituto “Mario Negri” Milano, 20154 Italy. <sup>2</sup>Department of Surgery and <sup>3</sup>Department of Oncology, Manzoni Hospital, Lecco 23900 Italy. <sup>4</sup>Clinic of Obstetrics and Gynaecology, University of Milano-Bicocca, San Gerardo Hospital, Monza, 20900 Italy. <sup>5</sup>Department of Biology, University of Padova, Padova, 35131 Italy. <sup>6</sup>Molecular Oncology Lab. IRCCS Candiolo, Torino, 10060 Italy. <sup>7</sup>Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy.

\*, equally contributed

#, corresponding authors

§ co-last.

**Keywords:** recurrent focal amplification; high grade serous ovarian cancer; multisite analysis

**Article category:** Research Article.

## Disclosure of Potential Conflicts of Interest

The authors declare no potential conflicts of interest

## Correspondence to:

Dr. Maurizio D’Incalci,

Department of Oncology IRCCS “Mario Negri” Institute for Pharmacological Research. Via Mario Negri 2, 20156 Milano Italy

+39-02-39014-473

maurizio.dincalci@marionegri.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.32288

### **List of abbreviations**

HGS-EOC: High Grade Serous Epithelial Ovarian Cancer

SNV: Single Nucleotide Variant

SCNA: Somatic Copy Number Alteration

TCGA: The Cancer Genome Atlas

FIGO: International Federation of Gynaecology and Obstetrics

WES: Whole Exome Sequencing

ddPCR: Droplet digital PCR

HR: Homologous Recombination

aCGH: Array CGH

FDR: False Discovery Rate

Accepted Article

## Abstract

High grade serous epithelial ovarian cancer (HGS-EOC) is a systemic disease, with marked intra and inter patient tumor heterogeneity. The issue of spatial and temporal heterogeneity has long been overlooked, hampering the possibility to identify those genomic alterations that persist, before and after therapy, in the genome of all tumor cells across the different anatomically districts. This knowledge is the first step to reveal those molecular determinants that characterize the tumor biology of HGS-EOC and their route towards malignancy. In this study, *-omics* data were generated from 79 snap frozen matched tumor biopsies, withdrawn before and after chemotherapy from 24 HGS-EOC patients, gathered together from independent cohorts. The landscape of somatic copy number variations (SCNAs) depicted a more homogenous and stable genomic portrait than the single nucleotide variant profile. GISTIC analysis identified two focal and minimal common region of amplification (FMCR) in the cytoband 3q26.2 (called region  $\alpha$ , 193 kb long), and 8q24.3 (called region  $\beta$ , 495 kb long). Analysis in two external databases, confirmed regions  $\alpha$  and  $\beta$  are features of HGS-EOC. The *MECOM* gene is located in region  $\alpha$ , while 15 genes are located in region  $\beta$ . No functional data are still available for the genes located in the  $\beta$  region. In conclusion, results have identified for the first time two FMCR of amplification in HGS-EOC, unveiling potential biological role in the etiopathogenesis of HGS-EOC

## Novelty and Impact Statement

This study analysed spatial and temporal molecular heterogeneity in HGS-EOC, by dissecting the molecular architecture of primary and matched synchronous and metachronous lesions. Despite tumor heterogeneity, two recurrent regions of focal amplification (3q26.2 and 8q24.3) were identified as a common hallmark of HGS-EOC. This finding paves the way to unveil the functional role of the 16 genes that mapped in these two genomic regions and to correlate their functions with the etiopathogenesis of HGS-EOC.

## Introduction

High-grade serous epithelial ovarian cancer (HGS-EOC), the most common and lethal sub-type of ovarian cancers, is a systemic disease, with multiple metastatic lesions wide-spread within the abdominal cavity (*i.e.*, synchronous lesions). HGS-EOC is generally sensitive to first line platinum (Pt)-based chemotherapy, although more than 80% of patients relapse with a disease that progressively becomes Pt-resistant. They tend to die within 5 years from the diagnosis<sup>1</sup>. With the exception of bevacizumab and PARP inhibitors, there are currently no targeted biological therapies approved for the first line treatment of HGS-EOC, while second line therapeutic options are still empirically administered based on the extent of Pt-free interval. The survival odds for a patient with HGS-EOC have not markedly improved despite years of extensive biological research and clinical trials.. These clinical disappointing results can be explained firstly, by the difficulty to identify a targetable genetic lesion among the complex, instable, and heterogeneous genomic background of HGS-EOC<sup>2</sup>. Secondly, tumor biopsies of relapsed resistant patients (*i.e.*, metachronous lesions) are not routinely collected, thus there is a dearth of information about genomic changes that characterize relapsed disease, how these have been acquired, and most importantly, whether these could guide the choice of second line treatments. In fact, the current paradigms to identify driver genomic lesions for diagnostic and prognostic purposes in HGS-EOC are largely derived from genomic information gained at the time of diagnosis, about the biology of the tumor growing in the ovary only. The issue of intra-patient tumor heterogeneity (*i.e.* spatial and temporal heterogeneity) has long been overlooked, and thus we currently do not know whether: *i*) the molecular scenario connected with tumor relapse mirrors that obtained at primary surgery in the ovary, or in any other synchronous lesions. *ii*) Whether the molecular features of relapsed disease are *de novo* acquired, due to Pt-exposure, or are still present in the primary tumor lesions and selected to grow after chemotherapy.

Recently, some studies with small cohorts of patients tried to answer to these questions by analyzing matched synchronous and, whenever possible, metachronous lesions. The few concordant data obtained so far suggest that: *i*) the branching evolution pattern of HGS-EOC, sustains the marked genomic heterogeneity of single nucleotide variants (SNV) among synchronous lesions<sup>3-6</sup>. Most of these genomic abnormalities arise in a random fashion and are lesion specific. *ii*) Relapsed-molecular features are largely derived by selective pressure of already distinct genetic events

present in the primary tumor rather than *de novo* acquired<sup>4,7</sup>. These findings implies that a single tumor lesion, the ovary for example, cannot any longer be considered representative of the entire disease, nor being informative on the genomic features of relapsed disease. Otherwise, the identification of driver genetic lesions that are recurrent among the different lesions of the same patient and characterize the genome of relapsed resistant disease, requires the systematic analysis of synchronous and, whenever possible, metachronous lesions.

In this study, spatially divided and temporally different tumor biopsies of the same patient have been investigated to identify in the genome of primary tumor lesions those discrete genomic abnormalities that are shared by all synchronous lesions and characterize at least in part the genomic landscape of matched relapsed disease.

## **Material and methods**

### **Patient cohort**

The study population consists of snap frozen multisite tumor biopsies withdrawn from women with diagnosis of serous carcinoma of high grade (grade 2 or 3) and advanced stage (FIGO stage III/IV International Federation of Gynaecological and Obstetrics criteria). Samples were gathered together from two independent clinical centers (named from now onwards as cohort A and B), and routinely stored in the Pandora tissue bank collection. Details are in Supplementary Results Section 1. Briefly, cohort **A** consists of 27 biopsies recruited from seven HGS-EOC patients undergoing debulking surgery at the Division of Gynecology and Oncology Dept., Manzoni Hospital (Lecco, Italy) between September 2014 and March 2016. Cohort **B** consists of 52 biopsies withdrawn from 17 HGS-EOC patients undergoing primary and second surgery at the Obstetrics and Gynecology Dept., San Gerardo Hospital (Monza, Italy) between 1992 and 2010. To exclude potential confounding effects, fine pathological review was performed on matched paraffin blocks from each patient (Supplementary Results, Section 1). As previously reported, all biopsies selected for the study are characterized by tumor cell fraction >70%, with a comparable tumor percentage between matched synchronous and metachronous lesions<sup>3</sup>. The study was performed following the principles of the Declaration of Helsinki; the scientific ethical committees approved the collection and usage of tumor samples. Written informed consent was obtained from all patients. The results of this study are presented in accordance with reporting recommendations for tumor marker prognostic studies (REMARK) criteria, as detailed in Supplementary Results, Section 8<sup>8</sup>.

### **Whole Exome sequencing experiments and data analysis**

Genomic DNA from matched blood samples and snap frozen tumor biopsies was purified as detailed in Supplementary Methods Section 1. Whole Exome sequencing (WES) libraries were generated and analyzed as reported in Supplementary Methods, Section 2.

### **Array CGH experiments and data analysis**

Array CGH experiments were performed to identify allele specific absolute copy number alterations (SCNAs), as reported in Supplementary Methods, Section 4. Recurrent aberrations were detected with GISTIC2.0 software.

### **Digital droplet PCR**

Digital droplet PCR (ddPCR) experiments were performed in multiplex format as detailed Supplementary Methods, Section 6.

### **Statistical Analysis Methods**

A detailed list of all statistical analyses, the software used and the accompanying data are described in Supplementary Methods, Section 7.

### ***In silico* data validation**

Two sample cohorts were used for data validation. The first was a selection of cases (n= 494) available at The Cancer Genome Atlas (TCGA) consortium<sup>1</sup>. In particular from the 591 samples of serous ovarian cancer present in the repository, samples from patients with high grade Stage III and IV tumors and with copy number alteration data available were selected for the study (Supplementary Methods, Section 5). The second cohort was a selection of temporal and spatial biopsies with high data quality from a previously published data set (n=91)<sup>6</sup> from which segmented copy number data were available (see Supplementary Methods, Section 5).

### **Data availability**

Scripts, configuration files and full variant call sets are available at a repository on GitHub: <https://github.com/lbeltrame/mnegri-ov223>. Raw sequence files have been submitted to the European Genome-Phenome Archive (EGA) with accession number EGAS00001003048. Array CGH data files have been submitted to Array Express in accordance to the MIAME guidelines (ID E-MTAB-6900).

## Results

### Experimental design and cohort description

The aim of the study is to systematically capture the presence of those genomic aberrations that, beyond spatial and temporal heterogeneity, are recurrent among matched synchronous lesions of HGS-EOC patients, and persist in the tumor genome of relapsed disease, after chemotherapy exposure. To this end, genomic analysis was performed in a cohort of 79 snap frozen tumor biopsies obtained from 24 HGS-EOC patients, gathered together from two independent tumor tissue collections (cohort A and B, Figure 1). Analysis included samples from the ovary (across different anatomical regions, Table S1.1), omentum, fallopian tubes, peritoneal sites and other distant metastatic sites collected at the time of primary surgery (*i.e.*, chemotherapy naïve) and at relapse, after Pt-based chemotherapy, when the disease became progressively resistant to Pt-based therapy. Briefly, cohort A consists of 27 snap frozen synchronous tumor biopsies withdrawn from seven HGS-EOC patients at the time of primary surgery, before chemotherapy. Cohort B includes a heterogeneous group of 52 snap frozen tumor biopsies selected from 17 HGS-EOC patients and collected both at the time of primary surgery (*i.e.*, 30 synchronous lesions) and at relapse after one or more lines of chemotherapy (*i.e.*, 22 metachronous lesions, Fig.1). All patients enrolled in cohort B died for relapsed resistant diseases<sup>3,4</sup>. Results were then compared *in silico* with published genomic data<sup>2,6</sup> to find out whether the identified recurrent genomic alterations were an hallmark of the HGS-EOC subtype. Demographics and clinical features of patients recruited in cohort A are detailed in Supplementary Results Section 1, while for cohort B they were similar to prior reports<sup>3,4</sup>.

### Multiregional WES analysis.

Multiregional WES analysis (150x mean coverage) was applied to 27 spatially distinct cryopreserved tumor biopsies and their matched normal blood samples, enrolled within cohort A (Figure 1). The aim of this part of the study was to initially identify those genomic abnormalities (*i.e.*, SNV, indel, and SCNAs) that are shared among the different synchronous lesions, with a low intra-patient heterogeneity. For each tumor sample, a catalogue of somatic and germline aberrations was generated (see details in Supplementary Results, Section 2). SNVs and indels were assigned to three classes: *i*) core mutations, the same mutated locus is present in the primary tumor as well as in all matched synchronous lesions. *ii*) Concordant mutations, the same mutated locus is present in at least two biopsies from the same patient. *iii*) Private mutations, each mutated locus is unique to a single anatomical site. Data reported in Table 1 show that the vast majority (~ 60%) of SNVs and indels are private to each single tumor sample, with core mutations ranging between 0.2-25%, only. To determine the global pairwise differences between two biopsies of the same patient or between two distinct patients, analysis of dissimilarity was performed. The matrix reported in

Figure S2.2 shows that each synchronous lesion is dissimilar one to each other by a degree of 50%, with inter-patient heterogeneity rising to 100%. The five different synchronous lesions from patient LC0044, three of which were sampled in three different anatomical regions of the right ovary, are a paradigmatic example of intra patient heterogeneity of SNVs (Figure S2.3). Focusing on the distribution of the mutated loci affecting the most frequently mutated genes in HGS-EOC, accordingly to the TCGA database (*TP53*, *BRCA1*, *BRCA2*, *NF1*, and *CDK12*)<sup>2</sup>, no core mutations other than those in the *TP53* gene (*c.452C>G*) were identified (Supplementary Table S2.4.). A pathogenic deletion of 22 nucleotides in the *BRCA1* gene was private of LC0044-C sample, while the other synchronous lesions from the same patient harbor a wt *BRCA1* gene. In conclusion, multiregional analysis of SNVs confirmed the marked heterogeneity of SNV loci among patients, among the different tumor sites of the same patient, or among different anatomical regions of each single tumor mass (*e.g.*, the different LC0044 ovarian samples). Based on these results, we next focused our attention on the SCNA profile, as we considered the catalogue of SNVs and indels not suitable for the aim of the study.

### **Identification of recurrent somatic copy number changes.**

As HGS-EOC is characterized by a preponderance of SCNA<sup>9</sup>, we exploited WES data to identify regions of SCNAs. To reduce the impact of technical noise, very low coverage regions were excluded from the analysis and stringent cutoffs were used (Supplementary Methods, Section 2). A total of 3021 SCNAs were identified, ranging from 3 to 417 SCNAs per samples, with median of 92 SCNAs per sample. The length of selected SCNAs ranged in size 1.5 kb-129 Mb, spanning a median of 1.45Mb (Table S2.3). Figure S2.4 shows the SCNAs pattern distribution, with a median of 200kbp long copy-number bins. All genomic regions that are indicative of somatic gain/amplifications are shown in red, whereas those that are indicative of somatic loss/deletions are shown in blue. Visual inspection of Figure S2.4 shows that tumors from different sites of the same donor show remarkable similar pattern, as exemplified by the comparable scatters of copy number bins across the five different synchronous lesions from LC0044 patient (Supplementary Results Section 2, Figure S2.5). As we were interested in identifying recurrent regions of gain or loss, we next focused on those SCNA that occur in at least 25% of samples. The most frequent genomic segment in gain was observed in cytoband 8q24 (48.08%, 5.37 Mb length, Supplementary Table S2.5). Gains in the cytobands 3q (8.72Mb), 5p (9.9Mb), 7q (8.22Mb), and 14q (1.47Mb) were reported with mean frequency close to 30%. Loss on chromosomes 5q (10.9Mb), 6q (8.13 Mb), 8p (9.9Mb), 13q (14.2Mb), 16q (7.26Mb), and 18q (7Mb) was observed with a mean frequency ranging from 40% to 48%. Losses in the 1p (17.06Mb), 4p (10.34Mb), and 4q (14.66Mb) were observed with a mean frequency close to 30%. The frequency

distribution of previously identified recurrent SCNA segments across the entirety of human chromosomes is depicted in Figure 2A.

### **Identification of focal and recurrent minimal SCNA with statistical significance**

As the SCNA profile was considered worthy of detailed investigation, we next decided to map recurrent regions of SCNA across the entire genome (not only in the coding sequencing), by array-CGH (aCGH), an in house well established state of the art technique for genome-wide detection of copy number changes. To identify tumor-specific genomic alterations and exclude regions of potential germline copy number variation, aCGH profiles were normalized against matched blood samples (resolution down to 44 kb, Supplementary Methods, Section 4). Analysis initially focused on synchronous lesions gathered together from cohorts A and B (*i.e.*, 57 biopsies from 24 patients, Figure 1). Results are detailed in Supplementary Section 3. Briefly, frequency plots distribution of SCNA obtained in the two cohorts separately, depict the presence of different genomic regions in gain or loss being gains on chromosome 3 and 8 those more frequently observed (25% and 70%, respectively, Figure S3.3). Merging aCGH data from both cohort A and B, highlighted the presence of five genomic regions on cytobands 3q, 5p, 7q, 8q, and 14q that are in gain in at least 30% to 50% of synchronous biopsies (Table S3.1). With the exception cytoband 14q (1.47Mb), all the other selected genomic regions are wide, spanning from 5Mb to 8 Mb in length. The frequency plot of recurrent SCNA segments generated by aCGH, across the 57 biopsies is depicted in Figure 2B.

To the aim of the study, it was important to narrow down the size of these regions to a minimal common region of alteration, and assign to a given location a statistical significance. We used the GISTIC algorithm (Genomic Identification of Significant Targets in Cancer, Supplementary Methods, Section 4), to identify recurrently altered SCNAs. Briefly, in this method the average magnitude of copy number alteration is used as a score, and a permutation-based test is used for estimating statistical significance. The GISTIC- G-score plot reveals a relatively simple pattern of amplification (Figure 3A). Using a 75% confidence level and a 5% FDR cutoff as thresholds, we identified a total of 46 “Focal GISTIC Peaks”, 26 gains and 20 losses (Table S3.2 and Table S3.3). We termed these recurrent alterations as focal and minimal common regions (FMCR). With regard to amplifications, GISTIC selects significant peaks on four different chromosomes (Figure 3A and Table S3.2). On chromosome 3, a peak was observed on cytoband 3q29, spanning a genomic region of 947.2 kb, that shows gain in 48.5% of samples. A peak was observed on cytoband 7q36.3 (17.5 kb, 55.8% of samples). Three regions were selected on chromosome 8: one on cytoband 8q23.2 (5Mb, 51.7% of samples) and two on cytoband 8q42.3 (495 Kb, 79.4% of samples; 75.3kbp 76.4% of samples). On chromosome 20, a peak was counted on cytoband 20q12.32 (2.5 kb, 48% of samples). Other GISTIC peaks for loss were reported but in less than 50% of samples (Table S3.3).

As a corollary, we compared the frequency plots distribution of SCNAs obtained by aCGH with those by WES, to evaluate whether platform-specific effects might have significant impact on data analysis and their robustness. No statistical differences were observed (t-test  $p=0.84$ ), limiting the impact of platform biases. In conclusion, different genomic techniques (*i.e.*, technical validation) across two independent tumor tissue collections (*i.e.*, clinical validation), confirmed the existence of genomic alterations that are recurrent across spatially different lesions of different HGS-EOC patients.

### Identification of FMCR in both synchronous and metachronous lesions

To the aim of the project, longitudinal analysis is of utmost importance to verify whether identified FMCR patterns in synchronous lesions persist at relapse after Pt-based chemotherapy. Since visual inspection (Figure S3.5) highlights marked concordance across matched synchronous and metachronous lesions of the same donor, we used GISTIC algorithm to precisely identify FMCR with level of significance in biopsies enrolled within cohort B. GISTIC algorithm identified 34 FMCR, 21 regions in gain and 13 regions in loss (Figure. 3B, Table S3.4, confidence  $>75\%$ ). Three regions in gain were present in more than 50% of cases (32 out of 52 samples), one mapped on cytoband 3q26.2, while the other two regions mapped on cytoband 8q24.3 (Table S3.4). To finally increase the statistical power of our analysis and to confirm the presence of commonly altered regions, GISTIC algorithm was used on the entire cohort of 79 samples (Figure 1). Analysis by CGH approach confirmed with level of significance, the three FMCR of gain in 3q26.3 and 8q24.2 (Figure 4 and Table 2). For clarity, these three regions will be called from now onwards as region  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively. Region  $\alpha$ , mapped on cytoband 3q26.2 and was observed in almost 65.8% of samples, spanning 193kbp long and only one gene (*i.e.*, *MECOM*) mapped in this region. Region  $\beta$  and  $\gamma$ , mapped on two different regions of the cytoband 8q24.3. Region  $\beta$ , is 495kbp long and was observed on almost 82.93% of our cases. 15 genes are known to map in this region (Table 2). Region  $\gamma$  is 620kbp long, and was observed in almost 78.05% of cases, including 17 known genes (Table 2).

FMCR of amplification are generally believed to encode oncogenes that can drive and sustain cancer growth<sup>10,11</sup>. To initially evaluate the possible functional role of the previously identified FMCR, genes listed in Table 2 were compared to those reported in the cancer gene census project (<http://cancer.sanger.ac.uk/cancergenome/projects/census/>). *MECOM* was described as carrying somatic mutations in leukemias, while germline variants in *RECQL4* were tied to osteosarcoma and skin cancer. No gene among those present in the regions was yet associated with copy number amplification. Pathway annotation of these genes associated them with a varying range of biological pathways, from DNA binding and regulation of gene expression to immune system and cell signaling regulation (Supplementary Results, Section 4, Table S4.1). Finally, primer pairs were designed in the coding sequence of selected genes

(Supplementary Results, Section 5, Figure S5.1) to validate by digital droplet PCR (ddPCR) FMCR of amplification in the cytobands 3q26.2 and 8q24.3. Orthogonal validation results reported in Figure S5.2 show that ddPCR analysis largely overlaps data generated CGH approach, confirming the robustness of our findings.

### **FMCR of amplification as a feature of HGS-EOC**

FMCR of amplification in  $\alpha$ ,  $\beta$  and  $\gamma$  are a common genomic feature of our cohort and persist in the genome despite dynamic changes among synchronous and metachronous lesions. We next questioned whether above identified FMCR were a feature of HGS-EOC or an exclusive characteristic of our cohort of patients. SCNAs from 494 HGS samples part of the TCGA repository were analyzed (Supplementary Methods, Section 5). The frequency plots reported in Figure S6.1 show that the long arm of chromosomes 3 and 8 are in gain (red) in more than 75% HGS-EOC tumor biopsy of the ovary. Fine mapping of the FMCR showed that  $\alpha$ , and  $\beta$  regions were present in 93 and 83 % of the TCGA samples respectively (Supplementary Results, Section 6, Table S6.1). No data were available for region  $\gamma$ . As TCGA only includes ovary biopsies taken at the time of diagnosis, we used another independent data set including both spatial and temporal biopsies<sup>6</sup> to confirm our results. Analysis of recurrent alterations on this data set with GISTIC (Supplementary Methods, Section 5) confirmed that region  $\alpha$  and  $\beta$  are present in gain in almost 74% and 85%, in the analyzed samples, respectively (Supplementary Results Section 6, Figure S6.2; Supplementary Results Section 6, Table S6.2). These results confirm that identified FMCR  $\alpha$  and  $\beta$  are a feature of HGS-EOC.

### **Clonal origin of 3q26.2 and 8q24.3**

As previously reported by our group and others, branched evolution trajectories depict the relationship among the different lesions of the same patient (Figure S7.1)<sup>4,12</sup>. In this context, to infer information on the possible biological role of gain of genomic material in the  $\alpha$  and  $\beta$  regions, we investigated the proportion of tumor cells harboring the two selected FMCR of amplification in the overall tumor cell population. The hypothesis is that dynamic changes in the cellular prevalence of a selected SCNA, could mirror a selective growth advantage in the presence of therapy exposure or across different micro-environmental niches. We used TITAN, a probabilistic framework, to predict by WES data the proportion of tumor cells harboring the  $\alpha$  and  $\beta$  regions of FMCR (Supplementary Methods, Section 2). Results reported in table S7.1 showed that: *i*) the prevalence of tumor cells in the bulky tissue is comparable between primary and matched synchronous lesions, and close to 70% (median 73%). This finding is in line with preliminary pathological data (Supplementary Section 1). *ii*) Almost 70% of the samples are characterized by a single tumor cluster

(defined as groups of cells sharing common alterations). *iii*) In the vast majority of lesions (85%), the FMCR of amplification in the  $\alpha$  and  $\beta$  regions, can be detected in all tumor cell populations (Supplementary Table S7.1). A representative example (patient LC0050) is shown in Figure S7.2 and Figure S7.3. When we analyzed difference in the cellular prevalence of  $\alpha$  and  $\beta$  regions of FMCR, compared to the other SCNAs identified by GISTIC analysis, we did not observe any difference ( $p > 0.1$ ). The last finding suggested that SCNAs could be an ancestral event in the tumor cell population, during branching evolution towards malignancy. As a corollary of this study, we used PhyloWGS algorithms (Supplementary Methods, Section 2) to reconstruct the genotypes of the different subpopulations based on the measured variant allelic frequency of SNVs and population frequencies of SCNA of bulk tumor samples. LC0050, with three different synchronous biopsies and with gain in the  $\alpha$  and  $\beta$  region present in the 100% of identified clusters was used as a representative case (Table S7.1). Data reported in Figure 5 show that the same clonal cell population (clone 1, blue circle) is present in all synchronous lesions. Clone 1 is characterized at genomic level by SCNA and in particular of FMCR in the  $\alpha$  and  $\beta$  regions. The other five sub-clones (clones 2, 3, 4, and clone 5 with its own derived clone 6) were largely specific to each tumor lesion, ranging from approximately 20% to 50% of tumor cell population (Figure 5) and characterized by copy number neutral LOH or SNVs (Table S7.2). In conclusion this preliminary analysis confirmed that SCNA and in particular the two selected  $\alpha$  and  $\beta$  regions of FMCR, are at the root of tumor evolution of HGS-EOC and thus are conserved across distinct tumor biopsies of the same patient and thus are still present in the genome of relapsed disease, after platinum based chemotherapy.

## Discussion

The overall data generated in this study allowed to identify two FMCR, namely  $\alpha$  and  $\beta$ , that are in gain in all lesions of the same patient, at diagnosis as well as at relapse after chemotherapy. The presence of these two FMCR have been further confirmed in two external databases of HGS-EOC, thus  $\alpha$  and  $\beta$ , can be considered *as bona fide*, a molecular hallmark of HGS-EOC.

Before a detailed discussion on the main findings of this study, there are some general comments that are worthy to be addressed. Firstly, intra-patient tumor heterogeneity represents an important clinical challenge in the development of therapeutic protocols. This study emphasizes the importance of systematic analysis of distinct matched synchronous and whenever possible metachronous lesions for the identification, at diagnosis of prognostic biomarkers that could be used to treat relapsed resistant disease. Due to the marked genomic instability and molecular heterogeneity of HGS-EOC, analysis of the single tumor biopsy withdrawn from the ovary is no more exhaustive.

Secondly, the molecular portrait of relapsed resistant disease is still largely unknown and difficult to predict from genomic information at diagnosis ( *BEItrame Annals of Oncology*). Many different factors, related both to the tumor genomic as well as to the tumor microenvironment are expected to shape the molecular features of relapsed disease and to influence patient's clinical outcome. Results from this study suggest that there are genetic factors (*i.e.* regions  $\alpha$  and  $\beta$ ) that despite dynamic changes caused by intrinsic (genetic instability) and extrinsic (treatment, tumor micro-environment) perturbations, persist in all tumor cells of distinct lesions at diagnosis as well as at relapse. In few words, the genomic landscape of relapsed disease is populated by the emergence of pre-existing resistant clones rather than *de novo* evolution.

Thirdly, while primary and relapsed HGS-EOC lesions lacks actionable point mutations, the SCNAs profile exhibited a low level of intra patient heterogeneity between matched synchronous and metachronous lesions. SCNAs are emerging as a dominant driving feature in the development and progression of several epithelial cancers<sup>13</sup>, leading to the identification of cancer-causing genes eligible for therapeutic approaches<sup>14-17</sup>.

Recurrent pattern of DNA amplification in the distal chromosome 8q together with gain in the 3q is broad and complex and has been previously reported in many different human cancers<sup>18-20</sup>. However, due to their large size, identification of specific driver genes within these regions is problematic. Chromosomal 8q24 has been considered the most important susceptibility region for prostate, colon, breast, as well as for ovarian cancers. Genome wide association studies identified association of breast, prostate and colorectal cancers with variants within 600kb region of a longer 1.18Mb sequence that does not contain any known gene and thus often referred as a "gene desert"<sup>21-27</sup>. Gain in the 8q23-24 was more often seen in metastatic tumors than in non-metastatic ones indicating that amplification of genes in these regions appear to affect the metastatic

potential of tumor cells<sup>28</sup>. Functional evidences indicated that this risk region may act as a regulatory hub by physical interactions with several neighboring genes important for carcinogenesis and metastatic process such as the oncogene c-MYC (8q24.12-24.13), the tyrosine phosphatase PRL-3 gene (also known as PTP4A3, 8q24.3) and the long non coding RNA PVT1 (8q24.21). Similarly, amplification of the 3q26-q27 region has been reported as a frequent and early event in a number of other epithelial cancers in particular in squamous cell carcinomas of the cervix<sup>29</sup>, esophagus<sup>30</sup>, prostate<sup>31</sup> and in the head and neck cancers<sup>32</sup>. Clusters of putative oncogenes are known to map in this region, including TERC (3q26.2), PI3KCA (3q26.32), ZASC1 (3q26.33), SCCRO (3q26.3), TP63 (3q27)<sup>30,33-35</sup>. A recently identified oncogene PRKCI is known to map in this region and its overexpression, in particular in HGS-EOC, engenders an immunosuppressive tumor microenvironment with poor infiltrating cells<sup>36</sup>. Focusing on EOC, amplification of the 8q24 is one of the most important copy number variant in HGS-EOC<sup>2</sup>, frequently associated with BRCA1 mutations<sup>37</sup>. Gain in 3q26 is found in almost 70% of TCGA sample<sup>2</sup>. It is reasonable to hypothesize that FMCR in  $\alpha$  and  $\beta$  hold protein coding genes or regulatory genes, important for the pathogenesis of HGS-EOC. However, data reported in the literature, in particular in HGS-EOC, have failed to identify a clear expression signature associated to 8q24 or 3q26 amplification that can confer an evolutionary advantage to tumor cells in terms of progression and metastatic spread. One possible explanation is that focusing on the tumor lesion growing in the ovary, identified genomic regions span several megabases, thus hampering the possibility to select those genes that are merely co-amplified (“passenger genes”, that are up-regulated due to their position, but do not necessarily bring any functional advantage to the tumor cells) together with those neighboring genes that have crucial biological functions to the malignant phenotype. In the present study, the strategy we adopted to increase the number of tumor biopsies per patients have allowed to narrow down the extension of the two susceptible loci in the cytoband 8q24 and 3p26 to minimal genomic regions that are in length 193kbp for 3p26 and 2495kbp for 8q24.3 respectively. FMCR in these two regions can drive a “dosage effect” on 16 known genes, 15 genes on 8q24.3 and one (*MECOM*, also known as *MDS1* and *EVII* complex locus) on 3q26.3. With the exception of *MECOM*, the vast majority of these genes are not functionally characterized according to data available in literature. *MECOM* is reported to be associated with favorable prognosis and platinum free interval in EOC<sup>7</sup> while for the 15 genes located on chromosome 8q24 to our knowledge no data are available on HGS-EOC. As a corollary of the study, we investigated on the possible molecular mechanisms responsible for above selected amplifications. Exome sequencing data generated on cohort were queried for the presence of chromosomal breaks or translocations in the two selected regions, with no significant results (data not shown).

In conclusion, the use of matched synchronous and metachronous lesions have allowed pinpointing for the first time two common genomic regions of SCNA that characterize the vast majority of HGS-EOC and

probably affect the biology of tumor cells growing in the ovary and those spread to the abdominal cavity at the time of diagnosis or at relapse after chemotherapy. Further research needs to clarify whether in those regions there are driver genes of HGS-EOC that so far have escaped identification.

Accepted Article

**Acknowledgments** We would like to thank Professor Andreas Gescher (Leicester, UK) for critical revision and editing of the manuscript. We are grateful to “Cloud4CARE” project for providing computational resources for data analysis. We acknowledge: 1) the “Nerina and Mario Mattioli” Foundation for supporting the activity of Pandora tumor tissue collection. 2) The Italian Association for Cancer Research (AIRC IG: 15177 and IG: 19997 to S. Marchini; IG 17185 to C. Romualdi), and CARIPLO Foundation (Grant Number, 2015-0848 to L. Beltrame and E. Calura). 3) The “Alessandra Bono Foundation” for supporting young investigators fellowships, and the genomic infrastructure. LP is recipient of a fellow from Italian Association for Cancer Research (Fellow Number 20996). GS is recipient of a fellow from Italian Association for Cancer Research (Fellow Number 19684).

**Competing Interest** The authors declare no competing financial interest in relation to the work described.

## References

1. Cannistra SA. Cancer of the ovary. *N Engl J Med* 2004;351:2519–29.
2. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;474:609–615.
3. Beltrame L, Di Marino M, Fruscio R, Calura E, Chapman B, Clivio L, Sina F, Mele C, Iatropoulos P, Grassi T, Fotia V, Romualdi C, et al. Profiling cancer gene mutations in longitudinal epithelial ovarian cancer biopsies by targeted next-generation sequencing: a retrospective study. *Annals of oncology : official journal of the European Society for Medical Oncology* 2015;26:1363–1371.
4. Paracchini L, Mannarino L, Craparotta I, Romualdi C, Fruscio R, Grassi T, Fotia V, Caratti G, Perego P, Calura E, others. Regional and temporal heterogeneity of epithelial ovarian cancer tumor biopsies: implications for therapeutic strategies. *Oncotarget* 2016;5.
5. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, Rosner J, Shumansky K, Kalloger S, Senz J, Yang W, McConechy M, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol* 2013;231:21–34.
6. Macintyre G, Goranova TE, De Silva D, Ennis D, Piskorz AM, Eldridge M, Sie D, Lewsley L-A, Hanif A, Wilson C, Dowson S, Glasspool RM, et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet* 2018;50:1262–70.
7. Lambrechts S, Smeets D, Moisse M, Braicu EI, Vanderstichele A, Zhao H, Van Nieuwenhuysen E, Berns E, Sehouli J, Zeillinger R, Darb-Esfahani S, Cacsire Castillo-Tong D, et al. Genetic heterogeneity after first-line chemotherapy in high-grade serous ovarian cancer. *Eur J Cancer* 2016;53:51–64.
8. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, Diagnostics SS of the N-EWG on C. Reporting recommendations for tumor marker prognostic studies (REMARK). *Journal of the National Cancer Institute* 2005;97:1180–1184.
9. Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, Nones K, Cowin P, Alsop K, Bailey PJ, Kassahn KS, Newell F, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 2015;521:489–94.
10. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* 2011;12:R41.
11. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899–905.

- Accepted Article
12. Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, Gale D, Sayal K, Murtaza M, Baldwin PJ, Rosenfeld N, Earl HM, et al. Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis. *PLOS Medicine* 2015;12:e1001789.
  13. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–674.
  14. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhi R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;450:893–898.
  15. Eder AM, Sui X, Rosen DG, Nolden LK, Cheng KW, Lahad JP, Kango-Singh M, Lu KH, Warneke CL, Atkinson EN, Bedrosian I, Keyomarsi K, et al. Atypical PKC $\alpha$  contributes to poor prognosis through loss of apical-basal polarity and cyclin E overexpression in ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:12519–12524.
  16. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–1068.
  17. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, Golas B, Wang L, Motoi N, Szoke J, Reinersman JM, Major J, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* 2009;28:2773–2783.
  18. Douglas EJ, Fiegler H, Rowan A, Halford S, Bicknell DC, Bodmer W, Tomlinson IPM, Carter NP. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer research* 2004;64:4817–4825.
  19. Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstrahl EJ, Jenkins RB. Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. *Journal of the National Cancer Institute* 1999;91:1574–1580.
  20. Osterberg L, Levan K, Partheen K, Staaf J, Sundfeldt K, Horvath G. High-resolution genomic profiling of carboplatin resistance in early-stage epithelial ovarian carcinoma. *Cytogenet Genome Res* 2009;125:8–18.
  21. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nature genetics* 2007;39:638–644.
  22. Schumacher FR, Feigelson HS, Cox DG, Haiman CA, Albanes D, Buring J, Calle EE, Chanock SJ, Colditz GA, Diver WR, Dunning AM, Freedman ML, et al. A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer research* 2007;67:2951–2956.
  23. Zanke BW, Greenwood CMT, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, et al. Genome-wide association

scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature genetics* 2007;39:989–994.

24. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature genetics* 2007;39:631–637.
25. Zhang B, Jia W-H, Matsuda K, Kweon S-S, Matsuo K, Xiang Y-B, Shin A, Jee SH, Kim D-H, Cai Q, Long J, Shi J, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nature genetics* 2014;46:533–542.
26. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature genetics* 2010;42:504–507.
27. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, Widschwendter M, Vierkant RA, Larson MC, Kjaer SK, Birrer MJ, Berchuck A, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nature genetics* 2010;42:874–879.
28. Ghadimi BM, Grade M, Liersch T, Langer C, Siemer A, Füzesi L, Becker H. Gain of Chromosome 8q23–24 Is a Predictive Marker for Lymph Node Positivity in Colorectal Cancer. *Clin Cancer Res* 2003;9:1808–14.
29. Sugita M, Tanaka N, Davidson S, Sekiya S, Varella-Garcia M, West J, Drabkin HA, Gemmill RM. Molecular definition of a small amplification domain within 3q26 in tumors of cervix, ovary, and lung. *Cancer genetics and cytogenetics* 2000;117:9–18.
30. Imoto I, Yuki Y, Sonoda I, Ito T, Shimada Y, Imamura M, Inazawa J. Identification of ZASC1 encoding a Krüppel-like zinc finger protein as a novel target for 3q26 amplification in esophageal squamous cell carcinomas. *Cancer research* 2003;63:5691–5696.
31. Sattler HP, Lensch R, Rohde V, Zimmer E, Meese E, Bonkhoff H, Retz M, Zwergel T, Bex A, Stoeckle M, Wullich B. Novel amplification unit at chromosome 3q25-q27 in human prostate cancer. *The Prostate* 2000;45:207–215.
32. Singh B, Stoffel A, Gogineni S, Poluri A, Pfister DG, Shaha AR, Pathak A, Bosl G, Cordon-Cardo C, Shah JP, Rao PH. Amplification of the 3q26.3 locus is associated with progression to invasive cancer and is a negative prognostic factor in head and neck squamous cell carcinomas. *The American journal of pathology* 2002;161:365–371.
33. Massion PP, Taflan PM, Jamshedur Rahman SM, Yildiz P, Shyr Y, Edgerton ME, Westfall MD, Roberts JR, Pietenpol JA, Carbone DP, Gonzalez AL. Significance of p63 amplification and overexpression in lung cancer development and prognosis. *Cancer research* 2003;63:7113–7121.

34. Yokoi S, Yasui K, Iizasa T, Imoto I, Fujisawa T, Inazawa J. TERC identified as a probable target within the 3q26 amplicon that is detected frequently in non-small cell lung cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2003;9:4705–4713.
35. Estilo CL, O-Charoenrat P, Ngai I, Patel SG, Reddy PG, Dao S, Shaha AR, Kraus DH, Boyle JO, Wong RJ, Pfister DG, Huryh JM, et al. The role of novel oncogenes squamous cell carcinoma-related oncogene and phosphatidylinositol 3-kinase p110alpha in squamous cell carcinoma of the oral tongue. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2003;9:2300–2306.
36. Sarkar S, Bristow CA, Dey P, Rai K, Perets R, Ramirez-Cardenas A, Malasi S, Huang-Hobbs E, Haemmerle M, Wu SY, McGuire M, Protopopov A, et al. PRKCI promotes immune suppression in ovarian cancer. *Genes & development* 2017;31:1109–1121.
37. George J, Alsop K, Etemadmoghadam D, Hondow H, Mikeska T, Dobrovic A, deFazio A, Group AOCS, Smyth GK, Levine DA, Mitchell G, Bowtell DD. Nonequivalent gene expression and copy number alterations in high-grade serous ovarian cancers with BRCA1 and BRCA2 mutations. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013;19:3474–3484.

## FIGURES AND TABLES LEGEND

**Figure 1 Graphical representation of matched biopsies enrolled in the study.** Cohort A consists of seven HGS-EOC patients with multiple matched tumor biopsies collected at time of primary surgery in the ovary (turquoise circles), or in other anatomical sites (light blue circle). All biopsies were naïve to chemotherapy. Cohort B consists of 17 patients with 16 tumor biopsies collected in the ovary (turquoise circles) and 14 matched synchronous tumor biopsies (light blue circle). 22 relapsed matched metachronous lesions were available after one or more lines of platinum based chemotherapy (dark blue circles). Matched blood samples (red circles) were used as reference.

**Figure 2. Frequency plots distribution of SCNA at primary surgery.** Frequency plots distribution of broad copy number aberrations with at least one copy number change (either in gain or loss), obtained from 27 biopsies collected at primary surgery and enrolled within cohort A (**Panel A**, WES technique). Panel B depicts SCNA obtained across the 57 biopsies enrolled in cohort A plus B, at time of primary surgery (aCGH technique). Details are described in Supplementary Methods, Section 2. The 22 autosomal and the X chromosome are arranged horizontally along the x-axis, from largest (on left) to smallest, with “p” arms to the left. At each genomic location, the percentage of tumors that have an aberration is shown on the y-axis. Red indicates recurrent copy number gain; blue indicates recurrent copy number loss. Numerical data of the distributions can be found in Supplementary Results, Section 2 and 3, Tables S2.5 and S3.1, respectively.

**Figure 3. GISTIC analysis of statistically significant recurrent copy number alterations.** Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm was used on the 57 primary tumor biopsies gathered together from cohort A and B (panel A) or on the 52 matched synchronous and metachronous lesions enrolled within cohort B (panel B). GISTIC calculates the statistically significant prevalence of copy number aberrations in the sample population ( $q$ -value  $< 0.05$ ). Each point represents a copy number segment identified in the data set. Points are proportionately spaced and arranged in genome order from 1pter to Xqter. The y axis indicates the G-score for each region (see Supplementary Methods, section 3): the higher the score, the more frequent the region is in the sample population. Red indicates significant regions of copy number gain, and blue significant regions of copy number loss. Grey peaks denote regions that are not statistically significant.

**Figure 4. GISTIC analysis on synchronous and metachronous lesions**

Genomic Identification of Significant Targets in Cancer (GISTIC) analysis of copy number gains was performed on 79 matched tumor synchronous and metachronous lesions gathered together from cohort A

and B GISTIC calculates the statistically significant prevalence of copy number aberrations in the sample population (q-value < 0.05). Points are proportionately spaced and arranged in genome order from 1pter to Xqter. The y-axis indicates the G-score for each region (see Supplementary Methods, section 3): the higher the score, the more frequent the region is in the sample population. Red indicates significant regions of copy number gain, and blue significant regions of copy number loss. The bottom plots show the median copy number for the three sample cohorts over part of chromosome 3 (left) and chromosome 8 (right). The pink area highlights the two significantly recurrent gain regions on 3q26.2 and 8q24.3. The lines indicate the median copy number for each chromosomal position: blue lines show the trend from Cohort A, pink lines from the spatial biopsies of Cohort B, and green lines from temporal biopsies of Cohort B.

**Figure 5. Sub-clonal analysis for patient LC0050.** The top tree shows the evolution of the different sub-clonal populations in all biopsies from patient LC0050 as predicted by PhyloWGS (Supplementary Methods, Section 2). 0 indicates an arbitrary original founder population, while the other numbers show the different inferred cell populations and their relationships. The middle line plot shows the relative percentage of tumor cell fraction for each of the six identified populations (as shown in the color key) in each of the biopsies. The bottom plot shows a visual reconstruction for each biopsy of the various cell populations based on the PhyloWGS data.

**Table 1. Private, concordant and core mutations in patients from cohort A.** Table reports for each patient, the total number of SNV identified across all matched synchronous lesions. SNVs are then classified as: *i*) private variants (*i.e.*, exclusive to a single tumor biopsy). *ii*) concordant variants (*i.e.*, shared by at least one matched synchronous lesion). *iii*) core variants (*i.e.*, shared across all lesions). The terms refer to the single genomic locus. \*, Patient LC0047 has two biopsies available, so all concordant mutations are also core mutations.

**Table 2. Genomic features of identified FMCR.** Table summarises the main genomic characteristics of the three FMCR regions, namely  $\alpha$  and  $\beta$  and  $\gamma$ , identified by GISTIC algorithm on the entire cohort of 79 snap frozen tumour biopsies, gathered together from cohort A and B. Coordinates refer to the GRCh37/hg19 genome assembly.

Table I

Patient code	Somatic variants			
	Total	Private (%)	Concordant (%)	Core (%)
LC008	721	475 (65.88)	195 (34.12)	51 (7.07)
LC0034	865	490 (56.71)	237 (27.43)	138 (16)
LC0037	454	305 (67.18)	104 (22.91)	45 (9.91)
LC0044	475	319 (67.16)	155 (32.63)	1 (0.21)
LC0047	321	239 (74.45)	0 (0)*	82 (25.55)
LC0050	442	278 (62.90)	60 (13.57)	104 (23.53)
LC0052	592	360 (60.81)	155 (26.18)	77 (13.02)

Table II

<b>FMCR</b>			
Acronym	$\alpha$	$\beta$	$\gamma$
Chromosome location	3q26.2	8q24.3	8q24.3
Type of alteration	Gain	Gain	Gain
Genomic position	168640383-168833772	144186874-144682631	145743041-146364022
Frequency in the sample cohort (%)	65.85	82.93	78.05
Approximate size (kb)	193	495	620
Number of named genes in region	1	15	17
Mapped genes	<i>MECOM</i>	<i>EEF1D, GLI4, LY6H, ZC3H3, C8orf51, GSDMD, ZNF696, TIGD5, NAPRT1, RHPN1, TOP1MT, ZFP41, GPIHBP1, MAFA, C8orf73</i>	<i>RPL8, ZNF7, ZNF16, RECQL4, LRRC14, COMMD5, ZNF250, C8orf33, ARHGAP39, ZNF34, ZNF251, ZNF252, TMED10P1, C8orf77, ZNF517, C8orf82, LRRC24</i>

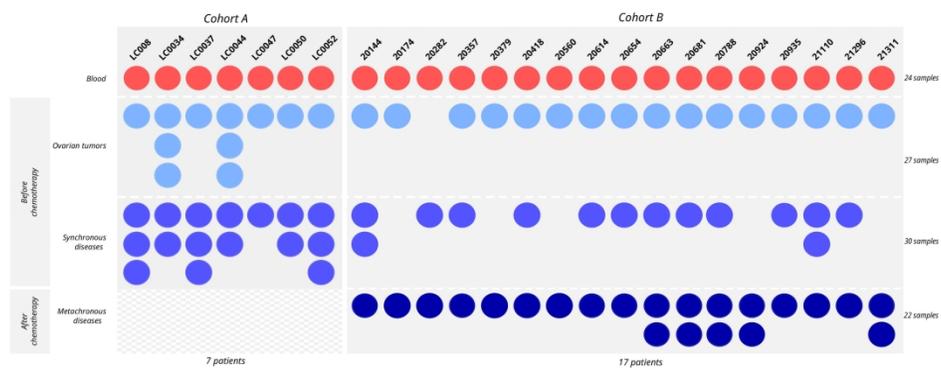


Figure 1

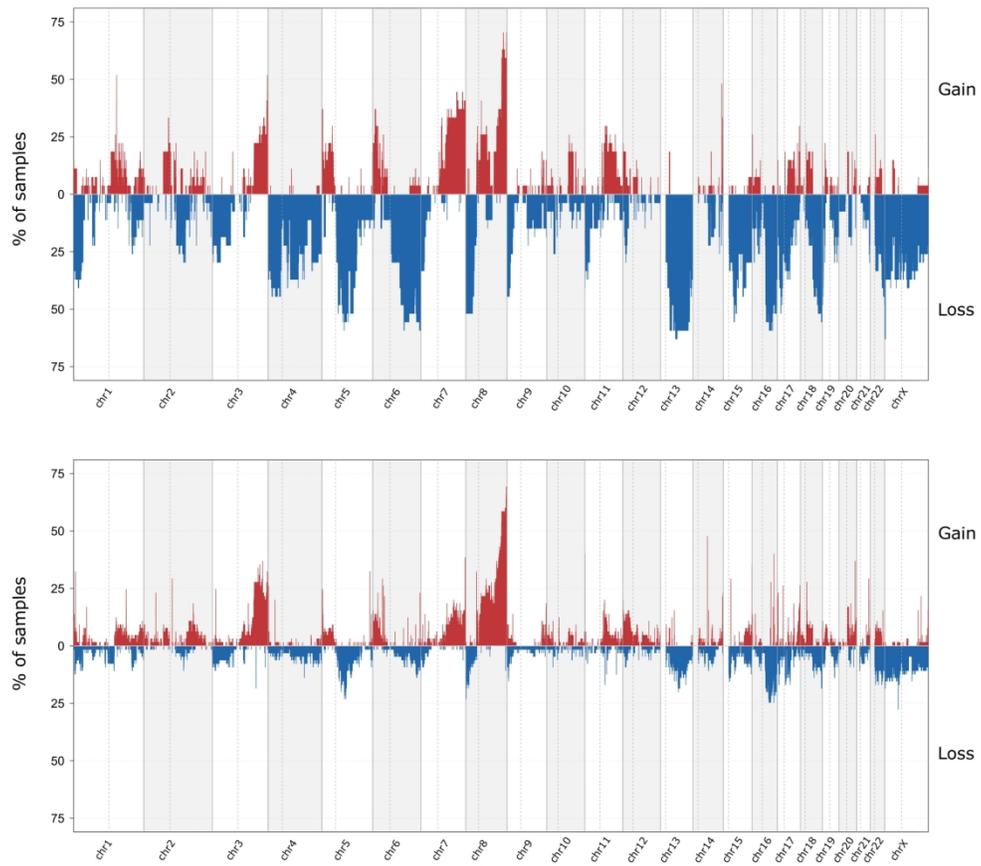


Figure 2

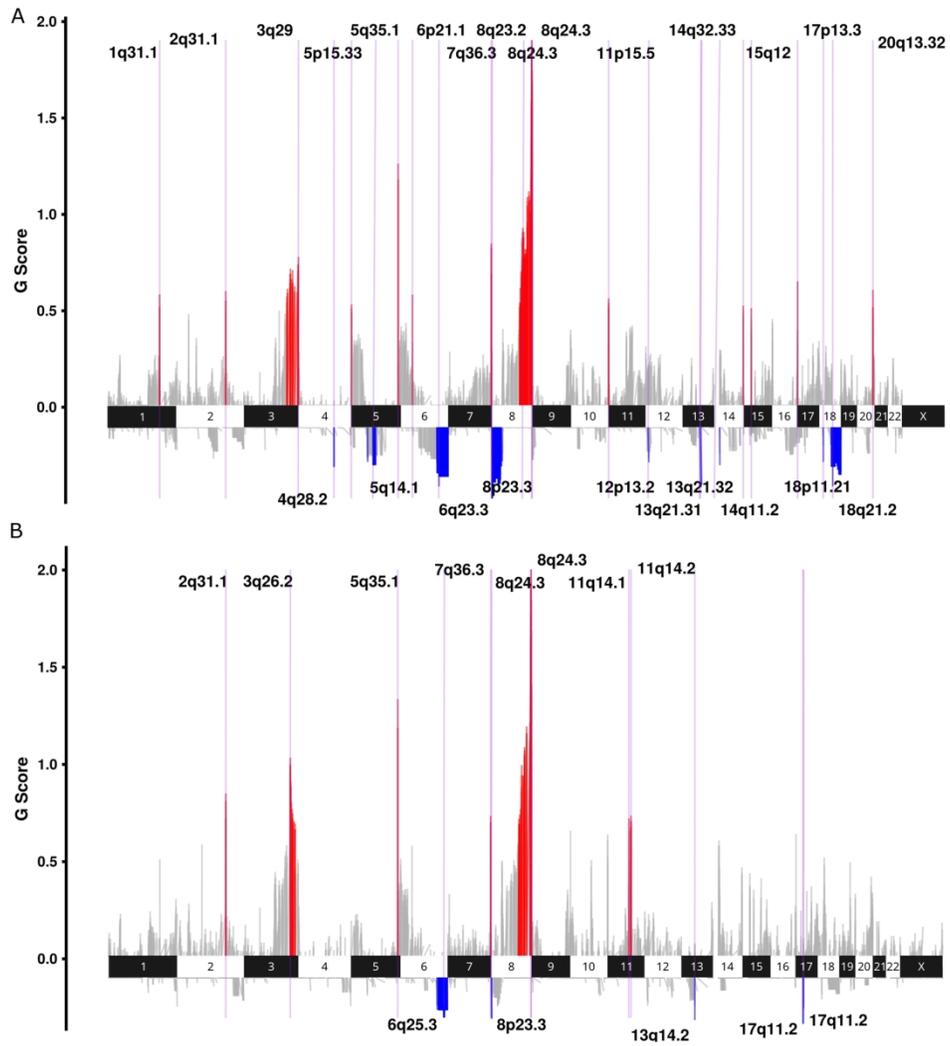


Figure 3

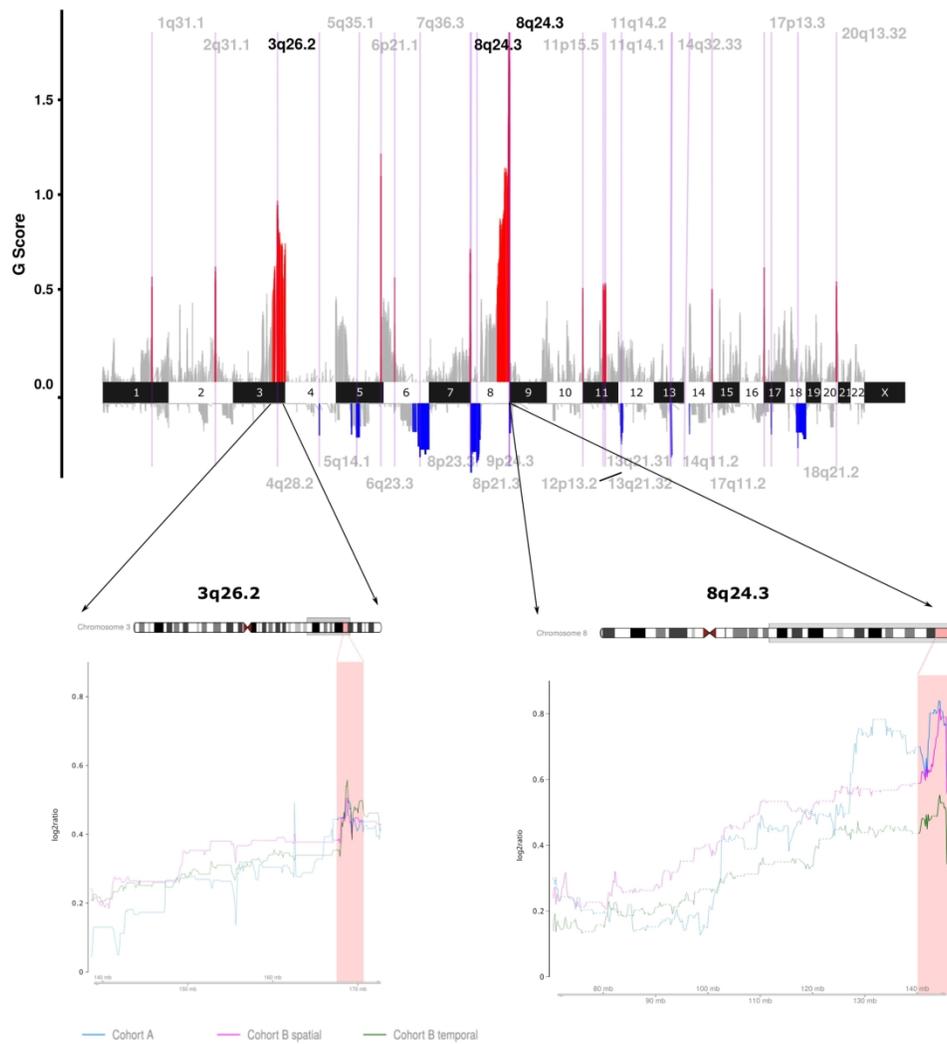


Figure 4

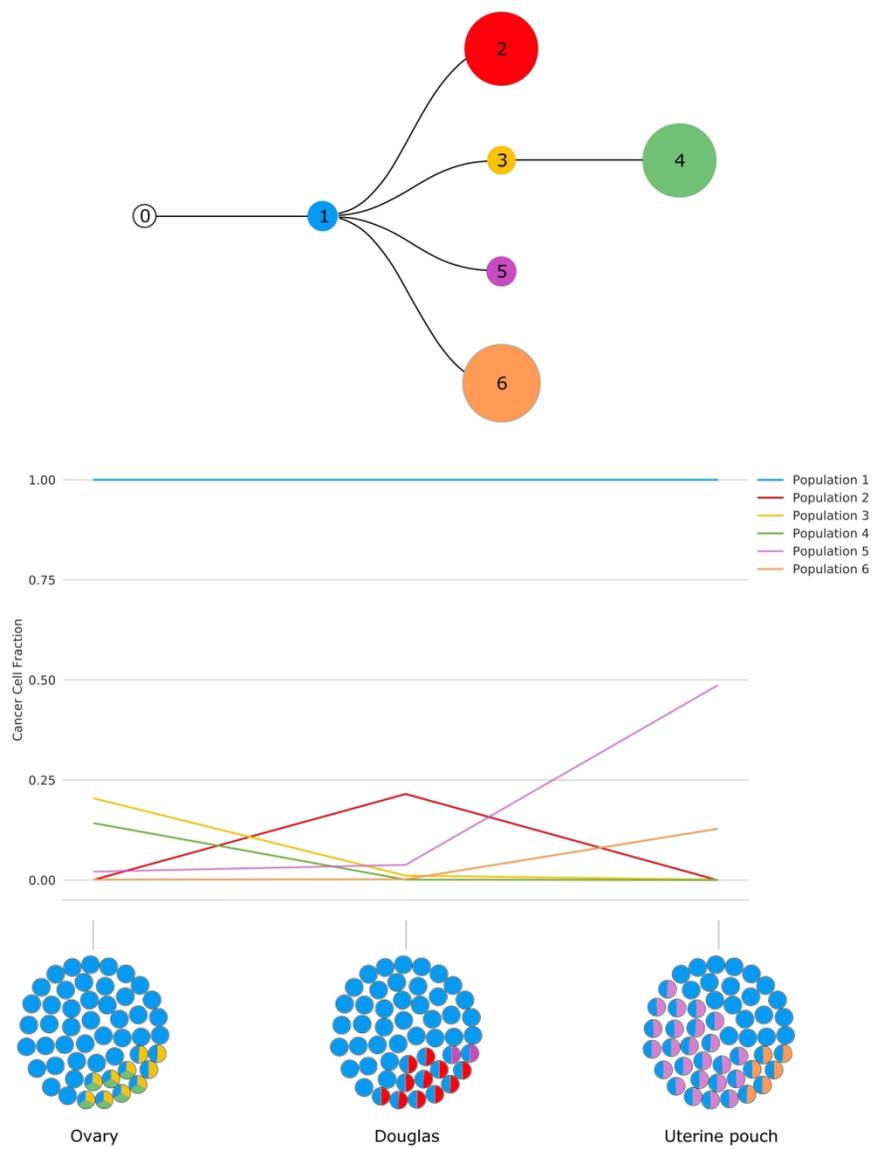


Figure 5