IJC
International Journal of Cancer

Multisite analysis of high-grade serous epithelial ovarian cancers identifies genomic regions of focal and recurrent copy number alteration in 3q26.2 and 8q24.3

Sara Ballabio[1*], Ilaria Craparotta[1*], Lara Paracchini[1*], Laura Mannarino[1], Silvia Corso[2], Maria Grazia Pezzotta[3], Martina Vescio[1,4], Robert Fruscio[5], Chiara Romualdi[6], Emanuele Dainese[3], Lorenzo Ceppi[5], Enrica Calura[6], Silvana Pileggi[1], Giulia Siravegna[7,8], Linda Pattini[4], Paolo Martini[6], Martina delle Marchette[5], Costantino Mangioni[2], Antonio Ardizzoia[3], Antonio Pellegrino[2], Fabio Landoni[5], Maurizio D'Incalci[1], Luca Beltrame[1†] and Sergio Marchini[1†]

[1]Department of Oncology, Istituto di Ricerche Farmacologiche "Mario Negri" IRCCS, Milano, Italy
[2]Department of Surgery, Manzoni Hospital, Lecco, Italy
[3]Department of Oncology, Manzoni Hospital, Lecco, Italy
[4]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy
[5]Clinic of Obstetrics and Gynaecology, University of Milano-Bicocca, San Gerardo Hospital, Monza, Italy
[6]Department of Biology, University of Padova, Padova, Italy
[7]Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Torino, Italy
[8]Department of Oncology, University of Torino, Candiolo, Torino, Italy

High-grade serous epithelial ovarian cancer (HGS-EOC) is a systemic disease, with marked intra and interpatient 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 anatomical districts. This knowledge is the first step to clarify those molecular determinants that characterize the tumor biology of HGS-EOC and their route toward malignancy. In our study, *-omics* data were generated from 79 snap frozen matched tumor biopsies, withdrawn before and after chemotherapy from 24 HGS-EOC patients, gathered together from independent cohorts.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 alterations depicts a more homogenous and stable genomic portrait than the single nucleotide variant profile. Genomic identification of significant targets in cancer analysis identified two focal and minimal common regions (FMCRs) of amplification in the cytoband 3q26.2 (region α, 193 kb long) and 8q24.3 (region β, 495 kb long). Analysis in two external databases confirmed regions α and β are features of HGS-EOC. The *MECOM* gene is located in region α, and 15 genes are in region β. No functional data are yet available for the genes in the β region. In conclusion, we have identified for the first time two FMCRs of amplification in HGS-EOC, opening up a potential biological role in its etiopathogenesis.

Introduction

High-grade serous epithelial ovarian cancer (HGS-EOC), the most common and lethal subtype of ovarian cancers, is a systemic disease, with multiple metastatic lesions widespread within the abdominal cavity (i.e., synchronous lesions). HGS-EOC is generally sensitive to first line platinum (Pt)-based

Key words: recurrent focal amplification, high-grade serous ovarian cancer, multisite analysis

Abbreviations: aCGH: array comparative genomic hybridization; ddPCR: droplet digital PCR; FDR: false discovery rate; FIGO: International Federation of Gynaecology and Obstetrics; HGS-EOC: high-grade serous epithelial ovarian cancer; HR: homologous recombination; SCNA: somatic copy number alteration; SNV: single nucleotide variant; TCGA: The Cancer Genome Atlas; WES: whole exome sequencing

Additional Supporting Information may be found in the online version of this article.

Conflicts of interest: The authors declare no potential conflicts of interest.

*S.B., I.C. and L.P. contributed equally to this work

†L.B. and S.M. shared co-last authorship

Grant sponsor: Associazione Italiana per la Ricerca sul Cancro; **Grant numbers:** Fellowship Number 19684, Fellowship Number 20996, IG: 15177, IG: 17185, IG: 19997; **Grant sponsor:** Fondazione Cariplo; **Grant number:** 2015-0848; **Grant sponsor:** Alessandra Bono Foundation; **Grant sponsor:** "Nerina and Mario Mattioli" Foundation

DOI: 10.1002/ijc.32288

History: Received 14 Jan 2019; Accepted 27 Feb 2019; Online 20 Mar 2019.

Correspondence to: Maurizio D'Incalci, Department of Oncology, Istituto di Ricerche Farmacologiche "Mario Negri" IRCCS, Via La Masa 19, 20156 Milano, Italy, Tel.: +39-02-39014-473, E-mail: maurizio.dincalci@marionegri.it

What's new?

Genomic analysis of high-grade serous epithelial ovarian cancer (HGS-EOC) is currently based on primary tumor testing at the time of diagnosis but tumors often relapse, and metastasis within the abdominal cavity form almost instantaneously. Here the authors analyzed spatial and temporal molecular heterogeneity in HGS-EOC by examining not only the primary tumor but also synchronous and matched relapse lesions. Despite tumor heterogeneity, two recurrent regions of focal amplification (3q26.2 and 8q24.3) were identified. Defining the functional role of the 16 genes that map to these two genomic regions will bring new insight into the pathogenesis of HGS-EOC and possibly define new therapeutic targets for this deadly disease.

chemotherapy, although more than 80% of patients relapse with a disease that progressively becomes Pt-resistant. They usually die within five years from the diagnosis.[1] 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 selected on the basis of the 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 disappointing clinical results can be explained first, by the difficulty to identify a targetable genetic lesion in the complex, unstable and heterogeneous genomic background of HGS-EOC.[2] Second, tumor biopsies of relapsed resistant patients (i.e., metachronous lesions) are not routinely collected, so there is a dearth of information about the genomic changes that characterize relapsed disease, how these are acquired and- most importantly- whether they could guide the choice of second line treatments. Current paradigms to identify driver genomic lesions for diagnostic and prognostic purposes in HGS-EOC are largely based on genomic information gained at the time of diagnosis, about the biology of the tumor growing only in the ovary. The issue of intra-patient tumor heterogeneity (i.e., spatial and temporal heterogeneity) has long been overlooked, so we 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 lesion; (*ii*) the molecular features of relapsed disease are acquired *de novo* due to Pt-exposure, or are already present in the primary tumor lesions and selected to grow after chemotherapy.

Recent studies with small cohorts of patients have tried to answer these questions by analyzing matched synchronous and, whenever possible, metachronous lesions. The few concordant data so far suggest that: (*i*) the branching pattern of HGS-EOC evolution sustains the marked genomic heterogeneity of single nucleotide variants (SNVs) among synchronous lesions.[3–6] Most of these genomic abnormalities arise in a random fashion and are lesion-specific; (*ii*) relapsed-molecular features are largely derived from selective pressure of already distinct genetic events present in the primary tumor rather than acquired *de novo*.[4,7] These findings imply that a single tumor lesion- the ovary for example- cannot any longer be considered representative of the entire disease, nor it is 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 our study, we investigated spatially divided and temporally different tumor biopsies from the same patient to identify the discrete genomic abnormalities in the genome of the primary ovarian tumor lesion that are shared by all synchronous lesions and at least characterize partly the genomic landscape of matched relapsed disease.

Materials and Methods

Patient cohort

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

Whole exome sequencing experiments and data analysis

Genomic DNA from matched blood samples and snap frozen tumor biopsies was purified as detailed in Supporting Information Methods Section 1. Whole exome sequencing (WES)

libraries were generated and analyzed as reported in Supporting Information Methods, Section 2.

Array CGH experiments and data analysis

Array comparative genomic hybridization (aCGH) experiments were done to identify allele-specific absolute copy number alterations (SCNAs), as reported in Supporting Information Methods, Section 4. Recurrent aberrations were detected with GISTIC2.0 software.

Digital droplet PCR

Digital droplet PCR (ddPCR) experiments were run in multiplex format as detailed Supporting Information Methods, Section 6.

Statistical analysis methods

All statistical analyses, the software used and the accompanying data are described in Supporting Information Methods, Section 7.

In silico data validation

Two sample cohorts were used for data validation. The first was a selection of 494 cases available at The Cancer Genome Atlas (TCGA) consortium.[2] From the 591 samples of serous ovarian cancer in the repository, were selected samples from patients with HGS-EOC stage III and IV tumors and with copy number alteration data available (Supporting Information Methods, Section 5). The second cohort was a selection of 91 temporal and spatial biopsies with high data quality from a previously published data set[6] where segmented copy number data were available (see Supporting Information 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 with 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 which besides 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. To this end, genomic analysis was done on a cohort of 79 snap frozen tumor biopsies from 24 HGS-EOC patients, gathered together from two independent tumor tissue collections (cohort A and B, Fig. 1). Analysis included samples from the ovary (across different anatomical regions, Supporting Information Table S1.1), omentum, fallopian tubes, peritoneal sites and other distant metastatic sites collected at 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 from seven HGS-EOC

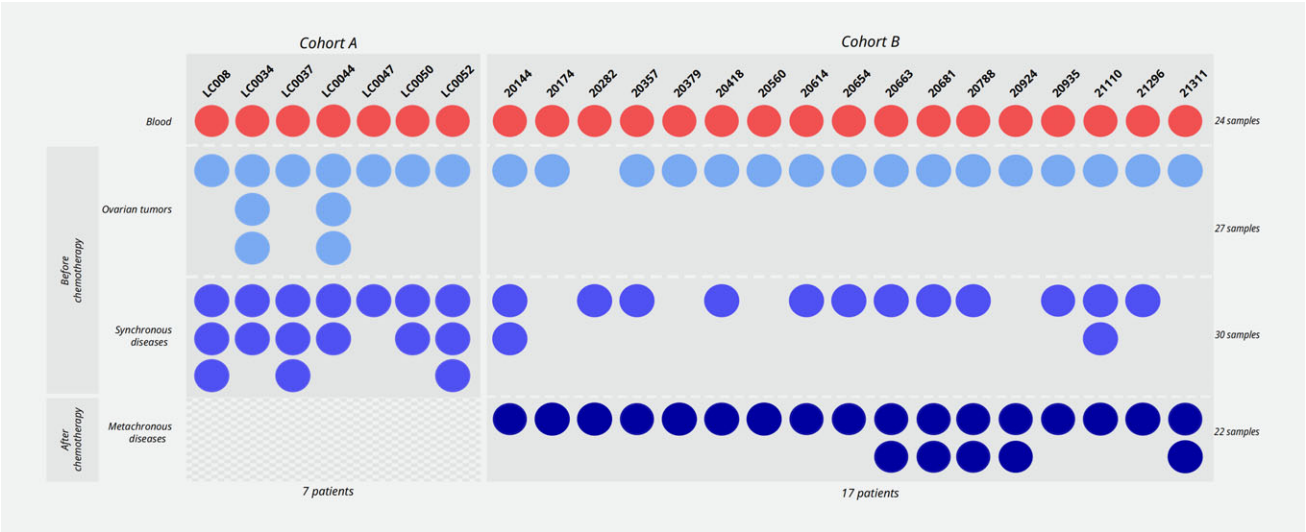

Figure 1. Graphical representation of matched biopsies. Cohort A consists of seven HGS-EOC patients with multiple matched tumor biopsies collected at time of primary surgery from the ovary (turquoise circles), or other anatomical sites (light blue circles). All biopsies were naïve to chemotherapy. Cohort B consists of 17 patients with 16 tumor biopsies from the ovary (turquoise circles) and 14 matched synchronous tumor biopsies (light blue circles). Twenty-two relapsed matched metachronous lesions were available after one or more lines of Pt-based chemotherapy (dark blue circles). Matched blood samples (red circles) were used as reference. [Color figure can be viewed at wileyonlinelibrary.com]

patients at primary surgery, before chemotherapy. Cohort B comprises a heterogeneous group of 52 snap frozen tumor biopsies selected from 17 HGS-EOC patients and collected both at 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 disease.[3,4] Results were then compared *in silico* with published genomic data[2,6] to find out whether the identified recurrent genomic alterations were hallmark of the HGS-EOC subtype. Demographics and clinical features of patients in cohort A are detailed in Supporting Information Results Section 1, while for cohort B they were similar to prior reports.[3,4]

Multiregional WES analysis

Multiregional WES analysis (150× mean coverage) was applied to 27 spatially distinct cryopreserved tumor biopsies and their matched normal blood samples in cohort A (Fig. 1). The aim of this part of the study was to initially identify the genomic abnormalities (i.e., SNVs, indels and SCNAs) that are shared among the different synchronous lesions, with low intrapatient heterogeneity. For each tumor sample, a catalog of somatic and germline aberrations was generated (for details see Supporting Information 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. Table 1 shows that the vast majority (~60%) of SNVs and indels are private to each single tumor sample, with core mutations ranging only between 0.2 and 25%. To determine the global pairwise differences between two biopsies of the same patient or between two distinct patients, we did an analysis of dissimilarity. The matrix reported in Supporting Information Figure S2.2 shows that each synchronous lesion is dissimilar to the others by 50%, with interpatient 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 (Supporting Information Fig. S2.3). Focusing on the distribution of the mutated loci affecting the most frequently mutated genes in HGS-EOC, according to the TCGA database (*TP53*, *BRCA1*, *BRCA2*, *NF1* and *CDK12*),[2] no core mutations were identified other than those in the *TP53* gene (*c.452C>G*) were identified (Supporting Information Table S2.4). A pathogenic deletion of 22 nucleotides in the *BRCA1* gene was private in LC0044-C sample, while the other synchronous lesions from the same patient harbored 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 catalog of SNVs and indels not pertinent to the aim of the study.

Table 1. Private, concordant and core mutations in patients from cohort A

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)[1]	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)

The table reports for each patient the total number of SNVs identified across all matched synchronous lesions. SNVs are then classified as (*i*) private variants (exclusive to a single tumor biopsy); (*ii*) concordant variants (shared by at least one matched synchronous lesion); (*iii*) core variants (shared across all lesions). The terms refer to the single genomic locus.

[1]Patient LC0047 had two biopsies available, so all concordant mutations are also core mutations.

Identification of recurrent somatic copy number changes

As HGS-EOC involves a preponderance of SCNAs,[9] 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 (Supporting Information Methods, Section 2). We identified 3,021 SCNAs ranging from 3 to 417 per samples, with median of 92 per sample. The length of the SCNAs ranged from 1.5 kb–129 Mb, with a median of 1.45 Mb (Supporting Information Table S2.3). Supporting Information Figure S2.4 shows the SCNAs pattern distribution, with a median of 200 kb long copy-number bins. All genomic regions indicative of somatic gain/amplification are shown in red, and those that are indicative of somatic loss/deletion are in blue. Supporting Information Figure S2.4 shows that tumors from different sites of the same donor show remarkably similar patterns, as exemplified by the comparable scatters of copy number bins across the five different synchronous lesions from patient LC0044 (Supporting Information Results Section 2 and Fig. S2.5). As we were interested in identifying recurrent regions of gain or loss, we next focused on the SCNAs that occured in at least 25% of samples. The most frequent genomic segment in gain was in cytoband 8q24 (48.08%, 5.37 Mb length, Supporting Information Table S2.5). Gain in the cytobands 3q (8.72 Mb), 5p (9.9 Mb), 7q (8.22 Mb) and 14q (1.47 Mb) were reported with mean frequency close to 30%. Loss on chromosomes 5q (10.9 Mb), 6q (8.13 Mb), 8p (9.9 Mb), 13q (14.2 Mb), 16q (7.26 Mb) and 18q (7 Mb) was observed with a mean frequency from 40% to 48%. Losses in the 1p (17.06 Mb), 4p (10.34 Mb) and 4q (14.66 Mb) 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 2*a*.

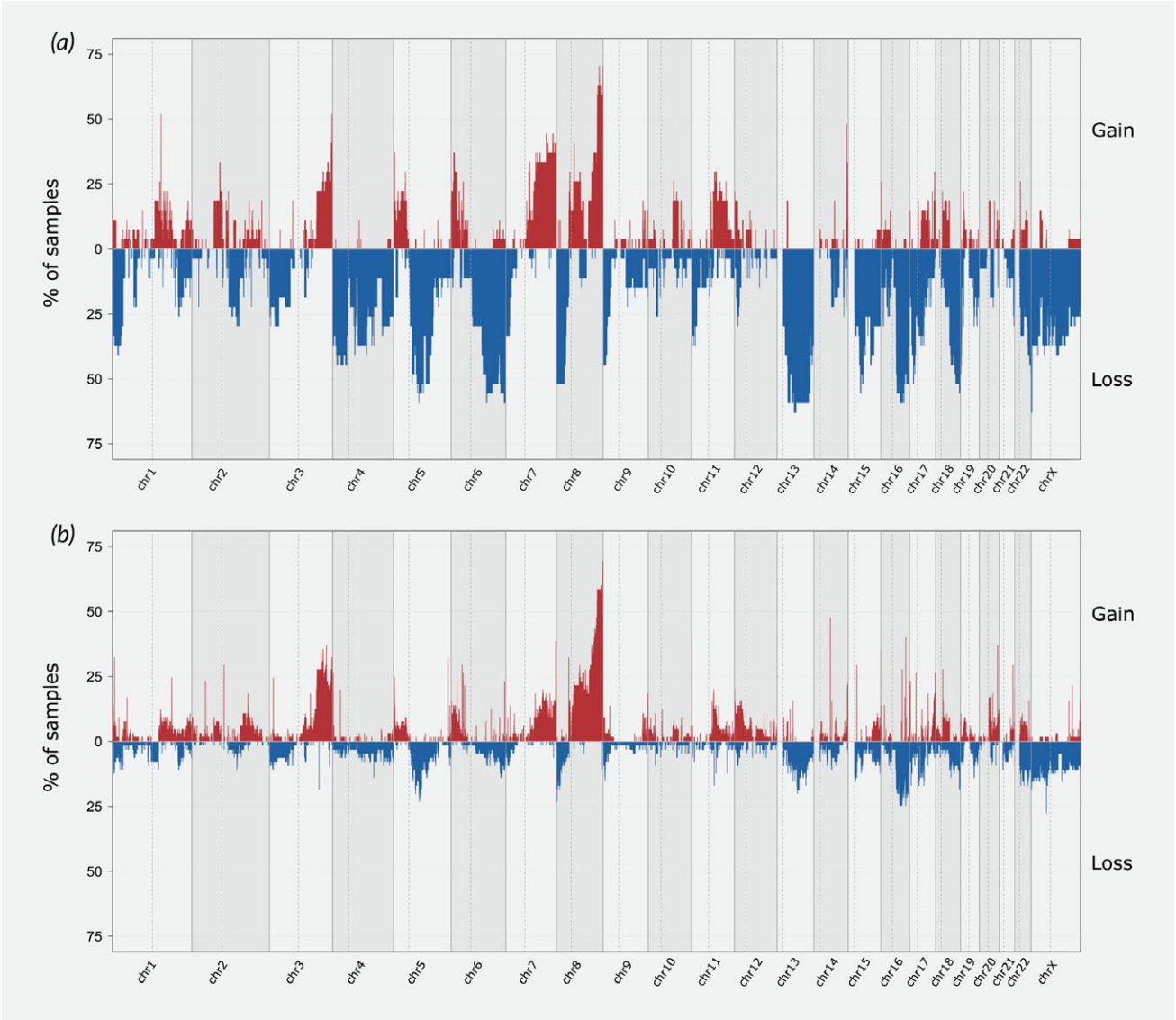

Figure 2. Frequency plots distribution of SCNAs at primary surgery. Frequency plots distribution of broad copy number aberrations with at least one copy number change (either gain or in loss) from 27 biopsies collected at primary surgery in cohort A (Panel *a*, WES technique). (Panel *b*) depicts SCNAs obtained across the 57 biopsies in cohort A plus B at time of primary surgery (aCGH technique). Details are given in Supporting Information 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 with an aberration is shown on the *y*-axis. Red indicates recurrent copy number gain and blue indicates recurrent copy number loss. Numerical data of the distributions can be found in Supporting Information Results, Section 2 and 3, Tables S2.5 and S3.1, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

Identification of focal and recurrent minimal SCNA with statistical significance

As the SCNAs profile was considered worthy of detailed investigation, we next mapped recurrent regions of SCNA across the entire genome (not only in the coding sequencing) by 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 changes, aCGH profiles were normalized against matched blood samples (resolution down to 44 kb, Supporting Information Methods, Section 4). Analysis initially focused on synchronous lesions gathered together from cohorts A and B (57 biopsies from 24 patients, Fig. 1). Results are detailed in Supporting Information Section 3. Briefly, frequency plots distribution of SCNAs obtained separately in the two cohorts, depicted the different genomic regions in gain or loss, with gain on chromosome 3 and 8 being the most frequent (25 and 70%, respectively, Supporting Information Fig. S3.3). Merging aCGH data from cohort A and B, highlighted the presence of ten genomic regions on cytobands 3q, 5q, 7q, 8p-8q,

11p, 14q, 15q, 16q and 20q, that show gain in at least 30 to 50% of synchronous biopsies (Supporting Information Table S3.1). The frequency plot of recurrent SCNA segments generated by aCGH across the 57 biopsies is depicted in Figure 2*b*.

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 a statistical significance to a given location. We used the genomic identification of significant targets in cancer (GISTIC) algorithm (Supporting Information Methods, Section 4), to identify recurrently altered SCNAs. Briefly, in this method the average magnitude of copy number alterations is used as a score, and a permutation-based test is used to establish a statistical significance. The GISTIC G-score plot reveals a relatively simple pattern of amplification (Fig. 3*a*). Using a 75% confidence level and a 5% FDR cutoff as thresholds, we identified 46 "Focal GISTIC Peaks", 26 gains and 20 losses (Supporting Information Table S3.2 and Table S3.3). We termed these recurrent alterations as focal and minimal common regions (FMCRs). With regard to amplifications, GISTIC selects significant peaks on four different chromosomes (Fig. 3*a* and Supporting Information Table S3.2). On chromosome 3, a peak was observed on cytoband 3q29, spanning a genomic region of 947.2 kb, that showed 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 (5.5 Mb, 51.7% of samples) and two on cytoband 8q42.3 (495 kb, 79.4% of samples; 75.3kb 76.4% of samples). On chromosome 20, a peak was counted on cytoband 20q13.32 (2.5 kb, 48% of samples). Other GISTIC peaks for loss were reported but in less than 50% of samples (Supporting Information Table S3.3).

As a corollary, we compared the frequency plot distribution of SCNAs obtained by aCGH with those by WES to see whether platform-specific effects had any significant impact on data analysis and their robustness. No statistical differences were observed (*t*-test $p = 0.84$), indicating a limited impact of platform biases.

In conclusion, different genomic techniques (technical validation) across two independent tumor tissue collections (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

The aim of the project implies that longitudinal analysis is of the utmost importance to verify whether FMCR patterns identified in synchronous lesions persist at relapse after Pt-based chemotherapy. Since visual inspection (Supporting Information Fig. S3.5) shows marked concordance across matched synchronous and metachronous lesions of the same donor, we used the GISTIC algorithm to precisely identify FMCR with the level of significance in biopsies in cohort B. The GISTIC algorithm identified 34 FMCR, 21 regions in gain and 13 in loss (Fig. 3*b*, Supporting Information 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 mapped on cytoband 8q24.3 (Supporting Information Table S3.4). Finally, to boost the statistical power of our analysis and confirm the presence of commonly altered regions, the GISTIC algorithm was used on the entire cohort of 79 samples. aCGH analysis confirmed the levels of significance of the three FMCR of gain in 3q26.3 and 8q24.2 (Fig. 4 and Table 2). For clarity, these three regions will be called regions α, β and γ, respectively. Region α, mapped on cytoband 3q26.2 and was observed in 65.8% of samples, spanning 193 kb long and only one gene (*MECOM*) mapped in this region. Regions β and γ mapped on two different regions of the cytoband 8q24.3. Region β is 495 kb long and was observed in almost 82.93% of our cases. Fifteen genes are known to map in this region (Table 2). Region γ is 620 kb long and was observed in almost 78.05% of cases, including 17 known genes (Table 2).

FMCRs of amplification are generally believed to encode oncogenes that can drive and sustain cancer growth.[10,11] To initially assess the possible functional role of the FMCRs identified, 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 in the regions has yet been 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 (Supporting Information Results, Section 4, Table S4.1). Finally, primer pairs were designed in the coding sequence of selected genes (Supporting Information Results, Section 5, Fig. 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 Supporting Information Figure S5.2 show that ddPCR analysis largely overlaps data generated by the aCGH approach, confirming the robustness of our findings.

FMCR of amplification as a feature of HGS-EOC

FMCR of amplification in α, β and γ 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 the FMCRs identified were a feature of HGS-EOC or exclusive to our cohort of patients. SCNAs from 494 HGS-EOC samples part of the TCGA repository were analyzed (Supporting Information Methods, Section 5). The frequency plots reported in Supporting Information 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 α and β regions were present in 93 and 83% of the TCGA samples respectively (Supporting Information Results, Section 6, Table S6.1). No data were available for region γ. As TCGA only includes ovary

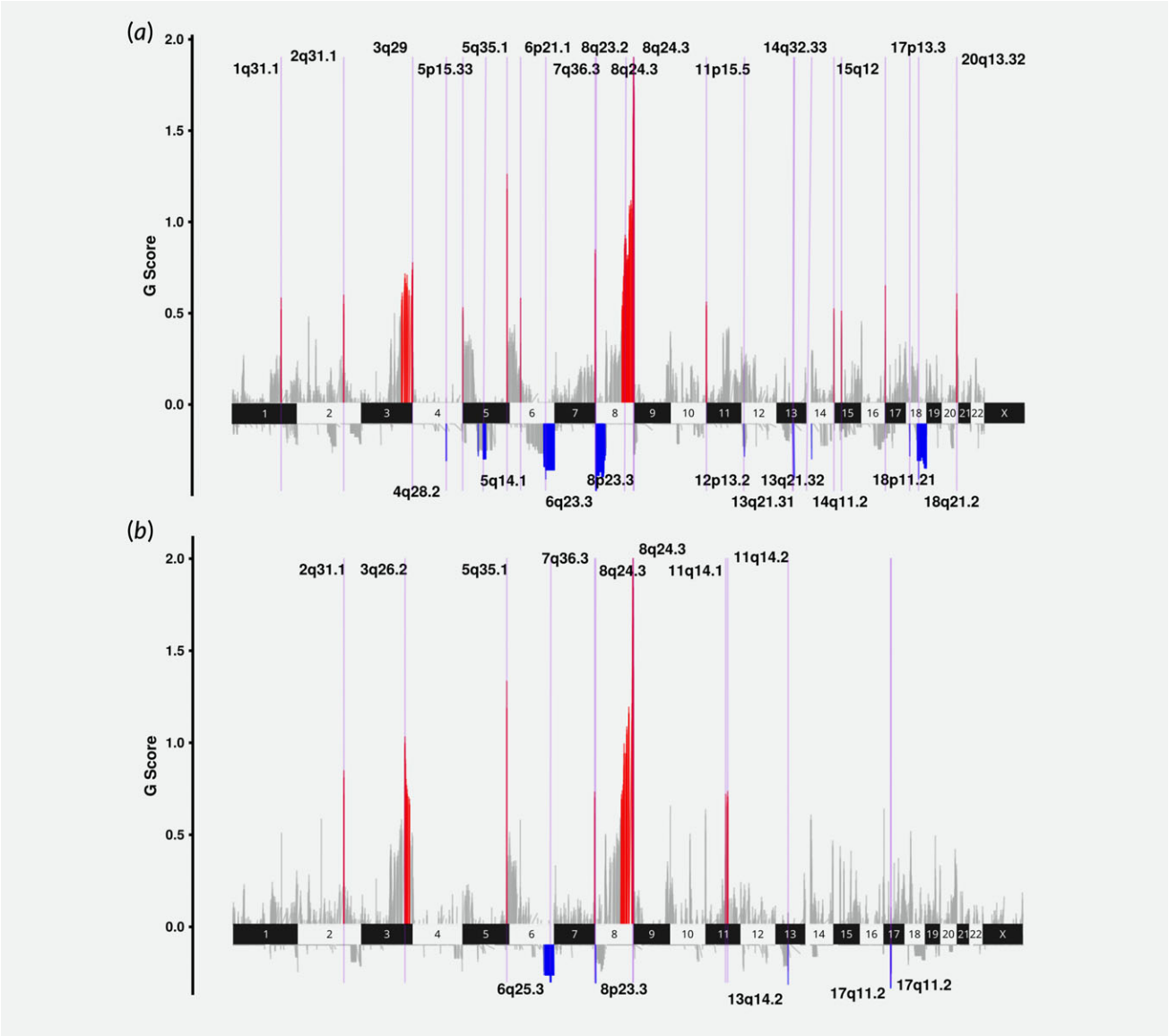

Figure 3. GISTIC analysis of statistically significant recurrent copy number alterations. The GISTIC algorithm was used on the 57 primary tumor biopsies from cohort A and B (panel a) or on the 52 matched synchronous and metachronous lesions in cohort B (panel *b*). GISTIC calculates the statistically 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 Supporting Information 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 indicates significant regions of copy number loss. Gray peaks denote regions that are not statistically significant. [Color figure can be viewed at wileyonlinelibrary.com]

biopsies taken at the time of diagnosis, we used another independent data set including both spatial and temporal biopsies[6] to confirm our results. Analysis of recurrent alterations on this dataset with GISTIC (Supporting Information Methods, Section 5) confirmed that regions α and β show gain respectively in almost 74 and 85% of the samples (Supporting Information Results Section 6, Fig. S6.2 and Table S6.2). These results confirm that FMCR α and β 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 (Supporting Information Fig. S7.1).[4,12] To obtain information on the possible biological role of gain of genomic material in the α and β regions, we investigated the proportion of tumor cells harboring the two selected FMCRs 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

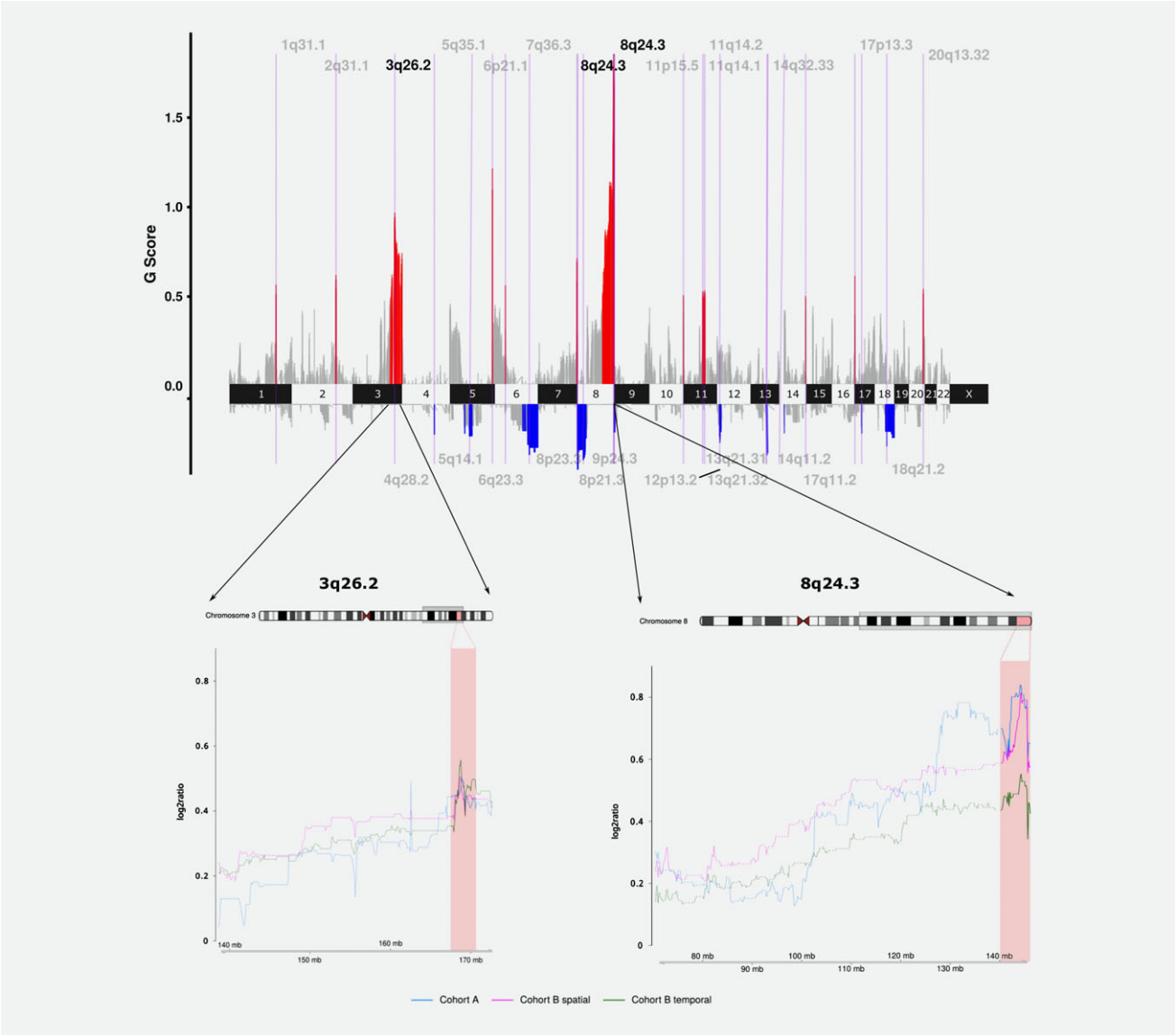

Figure 4. GISTIC analysis on synchronous and metachronous lesions. GISTIC analysis of copy number gains was done on 79 matched tumor synchronous and metachronous lesions from cohort A and B. GISTIC calculates the statistically significant prevalence of copy number aberrations in the sample population (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 Supporting Information 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 indicates 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. [Color figure can be viewed at wileyonlinelibrary.com]

growth advantage from therapeutic exposure or across different microenvironmental niches. We used TITAN, a probabilistic framework to predict from WES data the proportion of tumor cells harboring the α and β regions of FMCR (Supporting Information Methods, Section 2). The results reported in Supporting Information Table S7.1 showed that: (*i*) the prevalence of tumor cells in the bulky tissue is comparable in primary and matched synchronous lesions, and close to 70% (median 73%). This finding is in line with preliminary pathological data (Supporting Information Section 1). (*ii*) Almost 70% of the samples have 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 α and β regions can be detected in all tumor cell populations (Supporting Information Table S7.1). A representative example (patient LC0050) is shown in Supporting Information Figures S7.2 and S7.3. When we analyzed difference in the cellular prevalence of α and β regions of FMCR compared to the other SCNAs identified by GISTIC analysis, there was no difference

Table 2. Genomic features of identified FMCRs

FMCR			
Acronym	α	β	γ
Chromosome location	3q26.2	8q24.3	8q24.3
Type of alteration	Gain	Gain	Gain
Genomic position	168,640,383–168,833,772	144,186,874–144,682,631	145,743,041–146,364,022
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	*MECOM*	*EEF1D, GLI4, LY6H, ZC3H3, C8orf51, GSDMD, ZNF696, TIGD5, NAPRT1, RHPN1, TOP1MT, ZFP41, GPIHBP1, MAFA, C8orf73*	*RPL8, ZNF7, ZNF16, RECQL4, LRRC14, COMMD5, ZNF250, C8orf33, ARHGAP39, ZNF34, ZNF251, ZNF252, TMED10P1, C8orf77, ZNF517, C8orf82, LRRC24*

A summary of the main genomic characteristics of the three FMCR regions, namely α and β and γ, identified by GISTIC algorithm on the entire cohort of 79 snap frozen tumor biopsies, gathered together from cohort A and B. Coordinates refer to the GRCh37/hg19 genome assembly.

($p > 0.1$). This suggested that SCNAs could be an ancestral event in the tumor cell population, during branching evolution toward malignancy. As a corollary of our study, we used PhyloWGS algorithms (Supporting Information 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 α and β region present in the 100% of the identified clusters was used as a representative case (Supporting Information Table S7.1). Figure 5 shows that the same clonal cell population (clone 1, blue circle) was present in all synchronous lesions. Clone 1 is characterized at genomic level by SCNAs, particularly FMCR in the α and β regions. The other five subclones (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 the tumor cell population (Fig. 5) and characterized by copy number neutral LOH or SNVs (Supporting Information Table S7.2). In conclusion, this preliminary analysis confirmed that SCNAs and particularly the two selected α and β regions of FMCR are at the root of tumor evolution of HGS-EOC and are conserved across distinct tumor biopsies of the same patient.

Discussion

In this study we identified two FMCRs, namely α and β, that are in gain in all lesions in 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, so α and β, can be considered *as bona fide*, molecular hallmarks of HGS-EOC.

Before discussing on the main findings of this study, some general comments need to be addressed. First, intra patient tumor heterogeneity poses an important clinical challenge for the development of therapeutic protocols. Our 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 longer sufficient.

Second, the molecular portrait of relapsed resistant disease remains largely unknown and hard to predict from genomic information at diagnosis.[3] Many different factors related both to the tumor genome and its microenvironment are expected to shape the molecular features of relapsed disease and to influence the patient's clinical outcome. Results from our study suggest that there are genetic factors (i.e., regions α and β) that despite dynamic changes caused by intrinsic (genetic instability) and extrinsic (treatment, tumor microenvironment) perturbations persist in all tumor cells of distinct lesions at diagnosis as well as at relapse. Thus, the genomic landscape of relapsed disease is populated by the emergence of preexisting resistant clones more than *de novo* evolution.

Third, while primary and relapsed HGS-EOC lesions lack actionable point mutations, the SCNAs profile showed 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,[13] leading to the identification of cancer-causing genes eligible for therapeutic approaches.[14–17]

The recurrent pattern of DNA amplification in the distal chromosome 8q together with gain in the 3q is broad and complex and has been reported in many different human cancers.[18–20] However, their large size makes the identification of specific driver genes within these regions problematic. Chromosomal 8q24 has been considered the most important susceptibility region for prostate, colon, breast, and ovarian cancers. Genome wide association studies identified associations of breast, prostate and colorectal cancers with variants within 600 kb region of a longer 1.18 Mb sequence that does not contain any known gene and is thus often referred to as a

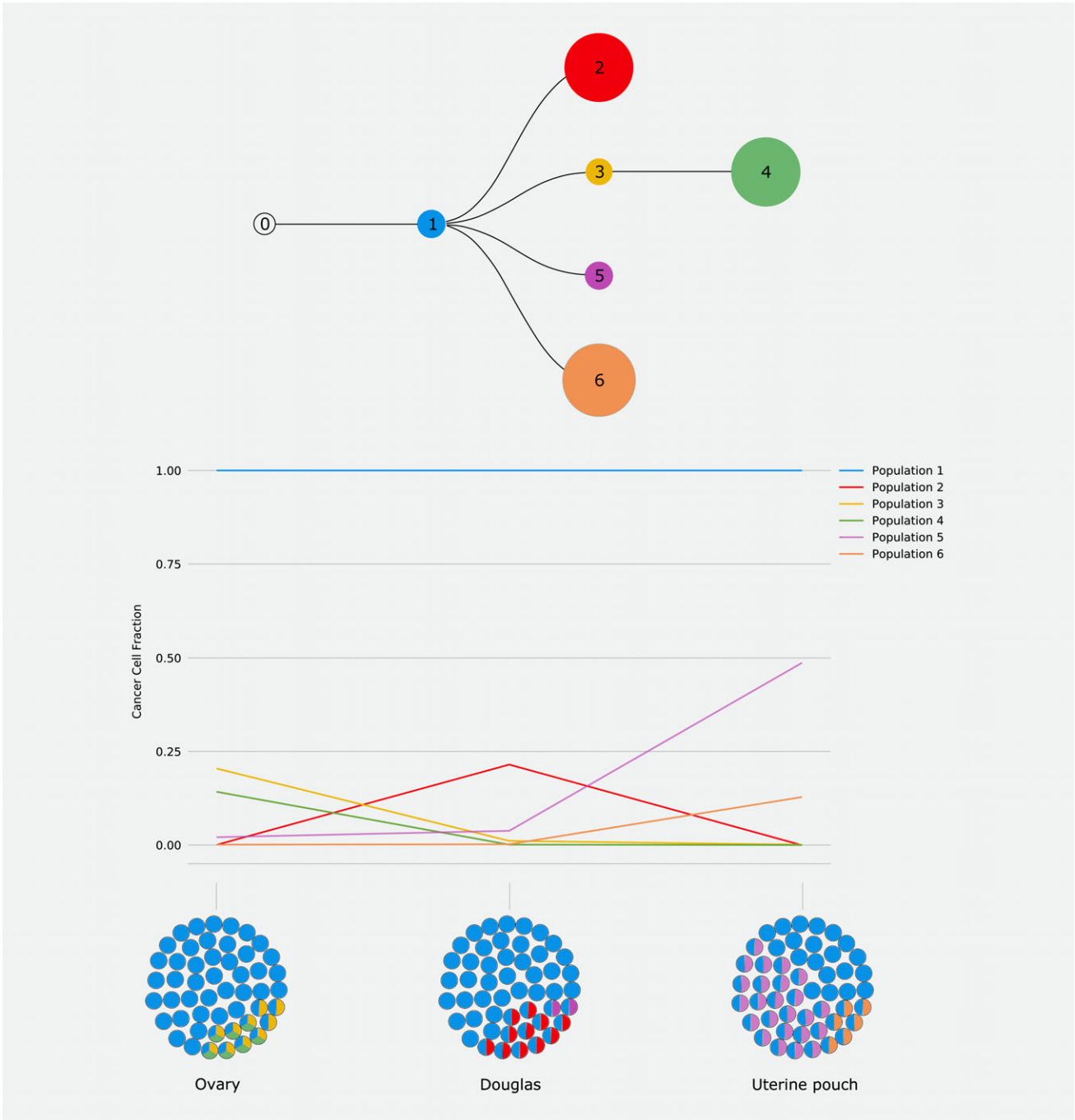

Figure 5. Subclonal analysis for patient LC0050. The top tree shows the evolution of the different subclonal populations in all biopsies from patient LC0050 as predicted by PhyloWGS (Supporting Information Methods, Section 2). Zero indicates an arbitrary original founder population, while the other numbers show the different inferred cell populations and their relationships. The middle line plot shows the 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. [Color figure can be viewed at wileyonlinelibrary.com]

gene desert.[21–27] Gain in the 8q23–24 was more often seen in metastatic tumors than non-metastatic ones, indicating that amplification of genes in these regions appear to affect the metastatic potential of tumor cells.[28] Functional evidences indicated that this risk region may act as a regulatory hub by physical interactions with several neighboring genes important for carcinogenesis and metastasis, 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 noncoding 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 particularly in squamous

cell carcinomas of the cervix,[29] esophagus,[30] prostate[31] and in the head and neck cancers.[32] Clusters of putative oncogenes map in this region, including *TERC* (3q26.2), *PIK3CA* (3q26.32), *ZASC1* (3q26.33), *SCCRO* (3q26.3) and *TP63* (3q27).[30,33–35] 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.[36] Focusing on EOC, amplification of the 8q24 is one of the most important SCNAs in HGS-EOC,[2] frequently associated with *BRCA1* mutations.[37] Gain in 3q26 is found in almost 70% of TCGA samples.[2] It is reasonable to hypothesize that FMCR in α and β 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 never identified any clear expression signature associated with 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, the identified genomic regions span several megabases, thus hampering the possibility to select those genes that are merely coamplified ("passenger genes", that are upregulated due to their position, but do not necessarily bring to the tumor cells any functional advantage) together with those neighboring genes that have crucial biological functions for 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 3q26 to minimal genomic regions that are 193 kb long for 3q26.2 and 2,495 kb long for 8q24.3. 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 *EVI1* complex locus) on 3q26.2. With the exception of *MECOM*, the vast majority of these genes are not functionally characterized according to the literature. MECOM is reported to be associated with favorable prognosis and Pt-free interval in EOC[7] while for the 15 genes located on chromosome 8q24, to the best of 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 did not reveal chromosomal breaks or translocations in the two selected regions (data not shown).

In conclusion, the use of matched synchronous and metachronous lesions served to pinpoint for the first time two common genomic regions of SCNAs that characterize the vast majority of HGS-EOC and probably affect the biology of tumor cells growing in the ovary and those spreading to the abdominal cavity at diagnosis or at relapse after chemotherapy. Further research needs to clarify whether there are driver genes of HGS-EOC in those regions that so far have escaped identification.

Acknowledgements

We would like to thank Professor Andreas Gescher (Leicester, UK) for critical revision and editing of the article. 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 SM; IG 17185 to CR). 3) The CARIPLO Foundation (Grant Number, 2015-0848 to LB and EC). 4) The "Alessandra Bono Foundation" for supporting young investigators fellowships, and the genomic infrastructure. 5) The SIA group for supporting young investigators fellowships. LP is the recipient of a fellow from Italian Association for Cancer Research (Fellowship Number 20996). GS is the recipient of a fellow from Italian Association for Cancer Research (Fellowship Number 19684). 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;474:609–15.
3. Beltrame L, Di Marino M, Fruscio R, et al. Profiling cancer gene mutations in longitudinal epithelial ovarian cancer biopsies by targeted next-generation sequencing: a retrospective study. *Ann Oncol* 2015;26:1363–71.
4. Paracchini L, Mannarino L, Craparotta I, et al. Regional and temporal heterogeneity of epithelial ovarian cancer tumor biopsies: implications for therapeutic strategies. *Oncotarget* 2016;5:1–14.
5. Bashashati A, Ha G, Tone A, 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, et al. Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet* 2018;50:1262–70.
7. Lambrechts S, Smeets D, Moisse M, 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, et al. Diagnostics SS of the N-EWG on C. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005;97:1180–4.
9. Patch A-M, Christie EL, Etemadmoghadam D, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 2015;521:489–94.
10. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;12:R41.
11. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899–905.
12. Schwarz RF, Ng CKY, Cooke SL, et al. Spatial and temporal heterogeneity in high-Grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med* 2015;12:e1001789.
13. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
14. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;450:893–8.
15. Eder AM, Sui X, Rosen DG, et al. Atypical PKCiota contributes to poor prognosis through loss of apical-basal polarity and cyclin E overexpression in ovarian cancer. *Proc Natl Acad Sci U S A* 2005;102:12519–24.
16. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–8.
17. Chitale D, Gong Y, Taylor BS, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* 2009;28:2773–83.
18. Douglas EJ, Fiegler H, Rowan A, et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 2004;64:4817–25.
19. Sato K, Qian J, Slezak JM, et al. Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. *J Natl Cancer Inst* 1999;91:1574–80.

20. Osterberg L, Levan K, Partheen K, et al. 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, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638–44.
22. Schumacher FR, Feigelson HS, Cox DG, et al. A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res* 2007;67:2951–6.
23. Zanke BW, Greenwood CMT, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;39:989–94.
24. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007;39:631–7.
25. Zhang B, Jia W-H, Matsuda K, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* 2014;46:533–42.
26. Turnbull C, Ahmed S, Morrison J, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010;42:504–7.
27. Goode EL, Chenevix-Trench G, Song H, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet* 2010;42:874–9.
28. Ghadimi BM, Grade M, Liersch T, et al. 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, et al. Molecular definition of a small amplification domain within 3q26 in tumors of cervix, ovary, and lung. *Cancer Genet Cytogenet* 2000;117:9–18.
30. Imoto I, Yuki Y, Sonoda I, et al. Identification of ZASC1 encoding a Krüppel-like zinc finger protein as a novel target for 3q26 amplification in esophageal squamous cell carcinomas. *Cancer Res* 2003;63:5691–6.
31. Sattler HP, Lensch R, Rohde V, et al. Novel amplification unit at chromosome 3q25-q27 in human prostate cancer. *Prostate* 2000;45:207–15.
32. Singh B, Stoffel A, Gogineni S, et al. 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. *Am J Pathol* 2002;161:365–71.
33. Massion PP, Taflan PM, Jamshedur Rahman SM, et al. Significance of p63 amplification and overexpression in lung cancer development and prognosis. *Cancer Res* 2003;63:7113–21.
34. Yokoi S, Yasui K, Iizasa T, et al. TERC identified as a probable target within the 3q26 amplicon that is detected frequently in non-small cell lung cancers. *Clin Cancer Res* 2003;9:4705–13.
35. Estilo CL, O-Charoenrat P, Ngai I, 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. *Clin Cancer Res* 2003;9:2300–6.
36. Sarkar S, Bristow CA, Dey P, et al. PRKCI promotes immune suppression in ovarian cancer. *Genes Dev* 2017;31:1109–21.
37. George J, Alsop K, Etemadmoghadam D, et al. Nonequivalent gene expression and copy number alterations in high-grade serous ovarian cancers with BRCA1 and BRCA2 mutations. *Clin Cancer Res* 2013;19:3474–84.