Vol. 9, No. 2, 2018
ISSN 2233-4203/ e-ISSN 2093-8950
LETTER www.msletters.org | Mass Spectrometry Letters

Loss of Potential Biomarker Proteins Associated with Abundant Proteins during Abundant Protein Removal in Sample Pretreatment

Jihoon Shin, Jinwook Lee, and Wonryeon Cho*

Department of Chemistry, Wonkwang University, 460 Iksandae-ro, Iksan, Jeonbuk 54538, Republic of Korea

Received May 15, 2018; Accepted May 17, 2018
First published on the web June 30, 2018; DOI: 10.5478/MSL.2018.9.2.51

Abstract : Capture of non-glycoproteins during lectin affinity chromatography is frequently observed, although it would seem to be anomalous. In actuality, lectin affinity chromatography works at post-translational modification (PTM) sites on a glycoprotein which is not involved in protein-protein interactions (PPIs). In this study, serial affinity column set (SACS) using lectins followed by proteomics methods was used to identify PPI mechanisms of captured proteins in human plasma. MetaCore, STRING, Ingenuity Pathway Analysis (IPA), and IntAct were individually used to elucidate the interactions of the identified abundant proteins and to obtain the corresponding interaction maps. The abundant non-glycoproteins were captured with the binding to the selected glycoproteins. Therefore, depletion process in sample pretreatment for abundant protein removal should be considered with more caution because it may lose precious disease-related low abundant proteins through PPIs of the removed abundant proteins in human plasma during the depletion process in biomarker discovery. Glycoproteins bearing specific glycans are frequently associated with cancer and can be specifically isolated by lectin affinity chromatography. Therefore, SACS using *Lycopersicon esculentum* lectin (LEL) can also be used to study disease interactomes.

Keywords : Serial affinity column set (SACS), Lectin, Depletion, Abundant protein removal, Protein-protein interactions (PPIs), Biomarker

Introduction

Affinity chromatography systems are becoming increasingly important as tools in the enrichment, identification, and quantification of proteins.[1-3] Quantification between disease and healthy samples using affinity chromatography reveals disease-related proteins in disease biomarker discovery.[4] However, a challenge in affinity chromatography is to capture affinity-selected proteins with reproducibility. In glycoproteomics, the great advantage of this affinity selection strategy is that immobilized lectins and antibodies allow specific glycan structures to be matched with particular proteins.[5-6] One of the objectives to use affinity chromatography is to obtain targeted glycoproteins reproducibly so that quantification between disease and healthy samples will reveal disease-related proteins. Disease-related proteins are frequently low abundant proteins which are easily suppressed by abundant proteins during mass analysis.[7-10] Therefore, biosamples such as plasma were depleted to remove those abundant proteins in the complex mixture.[11-14] The work being described here examines the utility of serial affinity column set (SACS) in understanding the interactions of abundant proteins with low abundant, disease-related and weakly bound proteins in protein complexes. Four serial agarose-bound *Lycopersicon esculentum* lectin (LEL) affinity columns, the LEL1 → LEL2 → LEL3 → LEL4 series, were examined in order to investigate protein-protein interactions (PPIs) among affinity selected proteins.[6]

*Reprint requests to Wonryeon Cho
E-mail: wcho@wku.ac.kr

Experimental

Materials and Chemicals

Agarose-bound LEL sorbent was purchased from Vector Laboratories (Burlingame, CA, U.S.A.). Normal pooled human plasma from 100 subjects was generously supplied by the National Institute of Standards and Technology (NIST, Gaithersburg, MD, U.S.A.). Acetic acid, sodium hydroxide, formic acid, calcium chloride, magnesium chloride, and HPLC grade acetonitrile (ACN) were purchased from Mallinckrodt Chemicals (Phillipsburg, NJ, U.S.A.). Ammonium bicarbonate, glycine, manganese chloride, proteomics grade N-p-tosyl-phenylalanine

chloromethyl ketone (TPCK)-treated trypsin, N-α-tosyl-l-lysine chloromethyl ketone hydrochloride (TLCK), 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES), iodoacetic acid (IAA), and l-cysteine were obtained from Sigma-Aldrich (St. Louis, MO, U.S.A.). Dithiothreitol (DTT), and urea were provided by Bio-Rad Laboratories (Hercules, CA, U.S.A.). PNGase F (glycerol free) was purchased from New England BioLabs (Ipswich, MA, U.S.A.). The C18 microspin column was obtained from The Nest Group, Inc. (Southborough, MA, U.S.A.). HLB Oasis SPE cartridges were provided by Waters (Milford, MA, U.S.A.). The DI water system was purchased from Millipore (Boston, MA, U.S.A.). Centrivap concentrator was purchased from Labconco, Corp. (Kansas City, MO, U.S.A.).

Serial Affinity Column Set (SACS)

Agarose-bound LEL sorbent was individually self-packed into four 4.6 mm × 50 mm columns. Healthy human plasma was loaded directly onto serially connected soft-gel affinity columns with mobile phase A (0.10 M HEPES buffer, pH 7.5 containing 1 mM $CaCl_2$ and 1 mM $MgCl_2$) at a flow rate of 0.3 mL/min. Following extensive washing with mobile phase A to remove nonspecifically and weakly bound proteins, the serial affinity columns were disassembled, and affinity-selected proteins were eluted from each column individually with solution B (0.5 M acetic acid–HCl solution, pH 2.5). All the elution curves were obtained with an absorbance detector operating at 280 nm using a 20 AD LC systems from Shimadzu Scientific Instruments, Inc. (Kyoto, Japan).

Saturation Test of Columns

A saturation test was done with an LEL column based on the previous papers.[5-6,15-16] The amount of injected sample was increased from 25 μg to 250 μg by 25 μg increments. The saturation amount of healthy human plasma samples was 200 μg of human plsma in the LEL column (data not shown). Therefore, the injection amount was decided as 200 μg of plasma below the saturation amount in order to prevent overloading.

Proteolysis

Captured proteins were desalted using HLB Oasis SPE cartridges and dried with a centrivap concentrator after being adjusted to pH 7.5 with a 0.5 M NH_4HCO_3 buffer and were then reconstituted with 8 M urea in a 50 mM HEPES buffer containing 10 mM $CaCl_2$. The denatured proteins were reduced with 10 mM DTT. After two hours of incubation at 50°C, iodoacetic acid was added to a final concentration of 20 mM for alkylation and incubated in darkness for an additional two hours. L-cysteine was then added to the reaction mixture to a final concentration of 40 mM, and the mixture was incubated for 30 minutes at room temperature. After dilution with a 50 mM HEPES buffer to a final urea concentration of 1 M, proteomics grade trypsin (2%, w/w, enzyme to protein) was added and incubated overnight at 37°C. The proteolysis reaction was stopped by the addition of TLCK (trypsin/TLCK ratio of 1:1 (w/w)). The resulting peptide mixture was desalted with HLB Oasis SPE cartridges and then dried into 50-100 μL with the centrivap concentrator, after being adjusted to pH 7.5 with the 0.5 M NH_4HCO_3 buffer.

PNGase F Digestion

N-Linked glycopeptides in a digested peptide mixture were deglycosylated by treatment with PNGase F. A 50 mM ammonium bicarbonate buffer was added to the desalted tryptic peptides to adjust to pH 7.0-8.0. Five microliters (2500 U) of PNGase F were added to the pH-adjusted tryptic peptides, and then the mixture was incubated overnight at 37°C. Following deglycosylation, samples were desalted using HLB Oasis SPE cartridges and concentrated using a C18 microspin column. The PNGase F-treated peptide mixtures were reconstituted in 0.1% formic acid solution and stored at -80°C until analysis with an LTQ-Orbitrap instrument.

LC-MS/MS-based Protein Identification

Proteins were identified with an LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher, Bremen, Germany). The peptide mixtures resulting from PNGase F treatment were separated on an Agilent 1100 HPLC system using a 75 μm × 120 mm C18 RPC column packed with 5 μm C18 Magic beads. Reversed phase chromatography (RPC) separations were achieved using a 60 min linear mobile phase gradient from 98% solvent A with 2% solvent B to 60% solvent A with 40% solvent B at a flow rate of 300 nL/min. Solvent A was composed of DI water to which formic acid had been added to a concentration of 0.1%. Solvent B was prepared with ACN to which formic acid had been added to a concentration of 0.1%. The electrospray ionization emitter tip was generated on the prepacked column with a laser puller (model P-2000, Sutter Instrument Co.). The HPLC system was coupled directly to the LTQ-Orbitrap hybrid mass spectrometer equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). The MS was operated in the data-dependent mode, in which a survey full scan MS spectrum (from *m/z* 300 to 1600) was acquired in the Orbitrap with a resolution of 60000 at *m/z* 400. This was then followed by MS/MS scans of the three most abundant ions with +2 to +3 charge states. Target ions already selected for MS/MS were dynamically excluded for 180 s. The resulting fragment ions were recorded in the linear ion trap.

Automated MS/MS data analysis was performed utilizing ProteinPilot software 5.0 with the Pro Group™ algorithm (ABI) for protein identification. The minimum acceptance criterion for peptide identification was a 99% confidence level. Most of the proteins were identified

based on the presence of at least two peptides from a protein identified by the Pro Group™ algorithm at the 99% confidence level. An unused score cutoff of 4 was the minimum value for identifying proteins with the ProteinPilot 5.0 Software.

PPI Analysis

MetaCore (Thomson Reuters, New York, NY, U.S.A.), STRING (www.string-db.org), Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA, U.S.A.), and IntAct (www.ebi.ac.uk/intact) were individually used to elucidate the interactions of the identified abundant proteins and to obtain the corresponding interaction maps.

The UniProtKB/Swiss-Prot accession number list for all 71 proteins in the SACS with LEL was first tabulated into the Excel spreadsheet, and this table was directly loaded into MetaCore for "batch search", which allowed simultaneous searches for all of the 71 proteins. The list of proteins corresponding to the loaded accession numbers was browsed; since one accession number might have multiple protein names, only protein names identical to the SACS list were selected and used for the preparation of the map for all the identified proteins using "build network" function. In option settings, only "direct" was selected whereas "indirect" was unselected; in the "additional options" settings, only "binding option" was selected whereas "low trust interactions", "functional interactions", and "use all compounds" were unselected. The individual interaction map of the abundant proteins in this experiment was identified and double checked through this process. The resulting maps were further corrected with other softwares in order of STRING, IPA, and IntAct.

Results and Discussion

A challenge with affinity chromatography is to capture affinity-selected proteins with reproducibility so that quantification between captured species may be verified. Especially in glycoproteomics, the great advantage of this affinity selection strategy is that immobilized lectins and antibodies allow specific glycan structures to be matched with a particular protein.[5-6] Thus, the serial affinity chromatography is a useful tool to validate targeted glycoproteins reproducibly so that quantification between disease and healthy samples will lead to the discovery of biomarkers for the disease.[5-6]

Capture of non-glycoproteins during lectin affinity chromatography is frequently observed, although it would seem to be anomalous. In reality, lectin affinity chromatography works at PTM sites on a glycoprotein which is not involved in PPIs. This paper explores the possibility that both types of interactions can occur simultaneously and be exploited in proteomic studies of the interactome. SACS using LEL columns was used to identify PPI mechanisms of captured proteins in human plasma.

Human plasma samples were injected into serially connected LEL columns. After extensive washing, the selected proteins in each column were eluted with an acidic mobile phase separately. Following serial affinity chromatographic selection, proteins selected in each of the 4 columns were trypsin digested, the peptide fragments separated by RPC, and fractions from RPC identified by tandem mass spectrometry. MetaCore, STRING, IPA, and IntAct were individually used to elucidate the interactions of the identified proteins and to obtain the corresponding interaction maps. Fifty-five glycoproteins were identified by SACS with LEL columns, and some of their proteoforms were resolved by displacement chromatography.[6]

It is intriguing that just connecting commonly-used affinity columns in series is useful in understanding displacement phenomena on affinity capture. The proteins captured by the LEL1 → LEL2 → LEL3 → LEL4 series are glycoproteins or proteins associated with the captured glycoproteins by PPIs. The 59 proteins captured by the LEL1 column in Table 1 either bore GlcNAc containing oligomers such as $(GlcNAc\beta_{1\text{-}4})_{1\text{-}4}$ or were associated with the glycoprotein(s) selected by the LEL1 column.[5]

Table 1 shows that human serum albumin (HSA) was the second most abundant protein among the identified proteins from the first LEL column (LEL1). However,

Table 1. Numbers of identified proteins and numbers of peptides used for the identification of HSA in the combinations of SACS; the LEL1 → LEL2 → LEL3 → LEL4 combination of SACS.

Selector	Numbers of Identified Proteins	HSA	
		Ranking[a]	Numbers of Peptides used for Protein Identificatiom
LEL1	59	2	38
LEL2	59	14	6
LEL3	54	16	4
LEL4	17	5	2

[a]Ranking means the order of identified proteins in each selector; ranking number 1 being the identified protein that used the most number of peptides and the number 59 (LEL1), 59 (LEL2), 54 (LEL3), and 17 (LEL4) being the protein with the least number of peptides used in each selector.

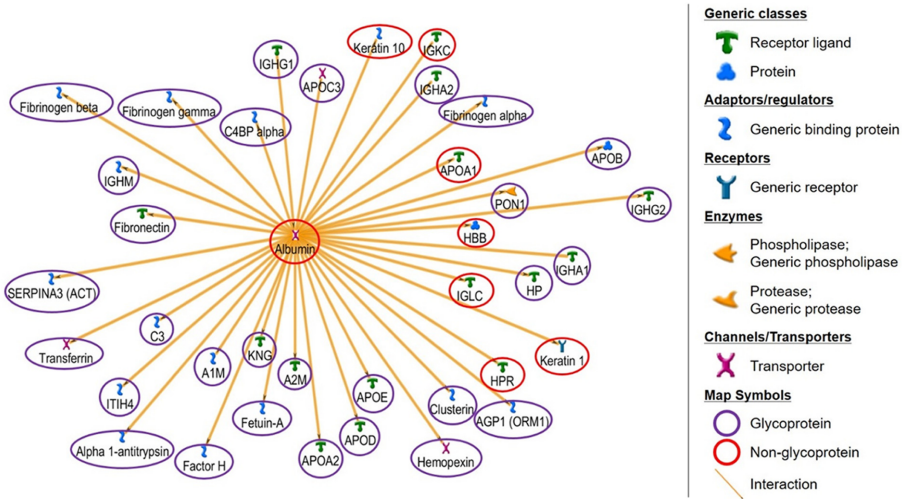

Figure 1. The biological network was generated using MetaCore mapping interactions between HSA and the other proteins in the identified total proteins of this experiment. Protein names for each symbols are as follows: Albumin, Human serum albumin; A1M, Protein AMBP; A2M, Alpha-2-macroglobulin; AGP1 (ORM1), Alpha-1-acid glycoprotein 1; Alpha 1-antitrypsin, Alpha-1-antitrypsin; APOA1, Apolipoprotein A-I; APOA2, Apolipoprotein A-II; APOB, Apolipoprotein B-100; APOC3, Apolipoprotein C-III; APOD, Apolipoprotein D; APOE, Apolipoprotein E; C3, Complement C3; C4BP alpha, C4b-binding protein alpha chain; Clusterin, Clusterin; Factor H, Complement factor H; Fetuin-A, Alpha-2-HS-glycoprotein; Fibrinogen alpha, Fibrinogen alpha chain; Fibrinogen beta, Fibrinogen beta chain; Fibrinogen gamma, Fibrinogen gamma chain; Fibronectin, Fibronectin; HBB, Hemoglobin subunit beta; Hemopexin, Hemopexin; HP, Haptoglobin; HPR, Haptoglobin-related protein; IGHA1, Ig lapha-1 chain C region; IGHA2, Ig alpha-2 chain C region; IGHG1, Ig gamma-1 chain C region; IGHG2, Ig gamma-2 chain C region; IGHM, Ig mu chain C region; IGKC, Ig kappa chain C region; IGLC, Ig lambda-1 chain C regions; ITIH4, Inter-alpha-trypsin inhibitor heavy chain H4; Keratin 1, Keratin, type II cytoskeletal 1; Keratin 10, Keratin, type I cytoskeletal 10; KNG, Kininogen-1; PON1, Serum paraoxonase /arylesterase 1; SERPINA3 (ACT), Alpha-1-antichymotrypsin; Transferrin, Transferrin.

HSA is a glycated protein, not a glycoprotein. Throughout the data analysis with four different softwares, HSA turned out to be captured as the second most abundant protein with many PPIs among glycoproteins and non-glycoproteins in this experiment as shown in Figure 1.

According to the map from the data analysis with multiple softwares, HSA was not captured by LEL column but captured by PPIs with 30 glycoproteins and seven non-glycoproteins such as Haptoglobin-related protein, Ig kappa chain C region, Keratin, type I cytoskeletal 10, Keratin, type II cytoskeletal 1, and Ig lambda-1 chain C regions in LEL column. HSA was identified from all four LEL columns with reasonable numbers of peptides as seen in Table 1. This can be proved by the fact that HSA was not the first rank among the identified proteins in the results, even though HSA is the most abundant protein in human plasma. It seems probable that most of HSA was bound to the first LEL column, LEL1, with PPIs of the bound proteins on LEL1, so the latter columns, LEL2, LEL3, and LEL4, have smaller amounts of HSA which resulted in lower rankings of HSA compared to LEL1 as described in Table 1. LEL4 also has a higher HSA ranking than LEL2 and LEL3 since the HSA abundance comes from PPIs with the bound proteins, not the non-specific binding on affinity selection. This identification method is one of the benefits of using SACS compared to a single affinity column.

Haptoglobin-related protein (HPR) is another abundant non-glycoprotein and was captured by PPIs with two glycoproteins and three non-glycoproteins through four LEL columns as elucidated in Figure 2. Ig gamma-1 chain C region (IGHG1) is also an abundant glycoprotein in plasma, which was captured with PPIs with three glycoproteins and nine non-glycoproteins as shown in Figure 3.

It looks like a flaw of structure analysis by SACS selection that there would be some false positives arising from proteins captured by the mechanisms such as PPIs other than the direct affinity selection of a targeted ligand. On the contrary, this mechanism elucidates PPIs among the captured glycoproteins by the affinity selector and the associated proteins with the glycoprotein(s).

The abundant protein removal may improve the identification capacity[10] of low abundant proteins, but it may lose precious disease-related low abundant proteins through PPIs of the removed abundant proteins in human plasma during the depletion process on biomarker discovery and quantification[17]. Therefore, depletion process in sample pretreatment for abundant protein removal should be considered with more caution.

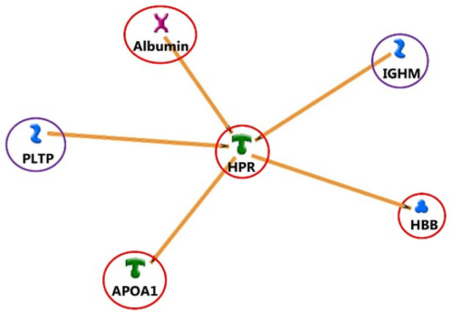

Figure 2. The biological network was generated using MetaCore mapping interactions between HPR and the other proteins in the identified total proteins of this experiment. Protein names for each symbols are as follows: Albumin, Human serum albumin; APOA1, Apolipoprotein A-I; HBB, Hemoglobin subunit beta; HPR, Haptoglobin-related protein; IGHM, Ig mu chain C region; PLTP, Phospholipid transfer protein.

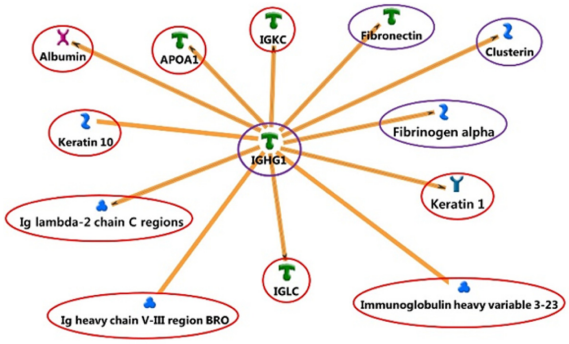

Figure 3. The biological network was generated using MetaCore mapping interactions between IGHG1 and the other proteins in the identified total proteins of this experiment. Protein names for each symbols are as follows: Albumin, Human serum albumin; APOA1, Apolipoprotein A-I; Clusterin, Clusterin; Fibrinogen alpha, Fibrinogen alpha chain; Fibronectin, Fibronectin; IGHG1, Ig gamma-1 chain C region; IGKC, Ig kappa chain C region; IGLC, Ig lambda-1 chain C regions; Ig heavy chain V-III region BRO, Ig heavy chain V-III region BRO; Ig lambda-2 chain C regions, Ig lambda-2 chain C regions; Immunoglobulin heavy variable 3-23, Ig heavy chain V-III region VH26; Keratin 1, Keratin, type II cytoskeletal 1; Keratin 10, Keratin, type I cytoskeletal 10.

Conclusions

SACS using LEL was used to identify PPI mechanisms of captured proteins. The abundant non-glycoproteins were captured with the binding to the selected glycoproteins. Therefore, depletion process in sample pretreatment for abundant protein removal should be considered with more caution because it may lose precious disease-related low abundant proteins through PPIs with the removed abundant proteins in human plasma during the depletion process on biomarker discovery. Glycoproteins bearing specific glycans are frequently associated with cancer and can be specifically isolated by SACS of affinity selectors such as lectins or antibodies. Furthermore, the results presented in this paper suggest that SACS of LEL can be used to study disease interactomes.

Acknowledgments

Authors acknowledge the support of Wonkwang University in 2012.

References

1. Azarkan, M.; Huet, J.; Baeyens-Volant, D.; Looze, Y.; Vandenbussche, G. *J. Chromatogr. B* **2007**, 849, 81.
2. Gadjeva, M. *The Complement System: Methods and Protocols*, Humana Press: Totowa, **2014.**
3. Gu, H.; Ren, J. M.; Jia, X.; Levy, T.; Rikova, K.; Yang, V.; Lee, K. A.; Stokes, M. P.; Silva, J. C. *Mol. Cell. Proteomics* **2016**, 15, 692.
4. Yang, N.; Feng, S.; Shedden, K.; Xie, X.; Liu, Y.; Rosser, C. J.; Lubman, D. M.; Goodison, S. *Clin. Cancer Res.* **2011**, 17, 3349.
5. Jung, K.; Cho, W. *Anal. Chem.* **2013**, 85, 7125.
6. Cho, W. *Anal. Chem.* **2015**, 87, 9612.
7. Gbormittah, F. O.; Hincapie, M.; Hancock, W. S. *Bioanalysis* **2014**, 6, 2537.
8. Kohler, J. J.; Patrie, S. M. *Mass Spectrometry of Glycoproteins: Methods and Protocols*, Humana Press: Totowa, **2013**.
9. Kim, B.; Araujo, R.; Howard, M.; Magni, R.; Liotta, L. A.; Luchini, A. *Expert Rev. Proteomics* **2018**, 15, 353.
10. Deutsch, O.; Krief, G.; Konttinen, Y. T.; Zaks, B.; Wong, D. T.; Aframian, D. J.; Palmon, A. *Rheumatology* **2015**, 54, 884.
11. Büyüktiryaki, S.; Uzun, L.; Denizli, A.; Say, R.; Ersöz, A. *Sep. Sci. Technol.* **2016**, 51, 2080.
12. Balmaña, M.; Sarrats, A.; Llop, E.; Barrabés, S.; Saldova, R.; Ferri, M. J.; Figueras, J.; Fort, E.; de Llorens, R.; Rudd, P. M.; Peracaula, R. *Clin. Chim. Acta* **2015**, 442, 56.
13. Mortezai, N.; Harder, S.; Schnabel, C.; Moors, E.; Gauly, M.; Schlüter, H.; Wagener, C.; Buck, F. *J. Proteome Res.* **2010**, 9, 6126.
14. Kang, U. -B.; Ahn, Y.; Lee, J. W.; Kim, Y. -H.; Kim, J.; Yu, M. -H.; Noh, D. -Y.; Lee, C. *BMC Cancer* **2010**, 10, 114.
15. Jung, K.; Cho, W.; Regnier, F. E. *J. Proteome Res.* **2009**, 8, 643.
16. Shin, J.; Cho, W. *Mass Spectrom. Lett.* **2016**, 7, 74.
17. Larkin, S. E. T.; Johnston, H. E.; Jackson, T. R.; Jamieson, D. G.; Roumeliotis, T. I.; Mockridge, C. I.; Michael, A.; Manousopoulou, A.; Papachristou, E. K.; Brown, M. D.; Clarke, N. W.; Pandha, H.; Aukim-Hastie, C. L.; Cragg, M. S.; Garbis, S. D.; Townsend, P. A. *Br. J. Cancer* **2016**, 115, 1078.