

University of Dundee

DOCTOR OF MEDICINE

The Role of Proteomics in Colorectal Cancer Biomarker Discovery

Goodbrand, Sarah A.

Award date:
2015

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CONTENTS

Declaration.....	1
Acknowledgements.....	2
Abbreviations.....	3
Abstract.....	6

INTRODUCTION

1.0 Background to Colorectal Cancer.....	7
1.1 Population Screening.....	11
1.2 Options for CRC Screening.....	12
1.2.1 Faecal Occult Blood Testing (FOBT).....	12
1.2.2 Flexible Sigmoidoscopy.....	17
1.2.3 Colonoscopy.....	18
1.2.4 CT Colonography.....	19
1.2.5 Biomarkers.....	20
1.2.5.1 Overview of Biomarkers.....	20
1.2.5.2 What makes a good Biomarker.....	21
1.2.5.3 Sources of Biomarkers.....	23
1.2.5.3.1 Tissue.....	23
1.2.5.3.2 Blood.....	24
1.2.5.3.3 Other Proximal Fluids.....	26
1.2.5.3.4 Cell Lines.....	27
1.2.5.3 Phases of Biomarker Discovery.....	29
1.2.6 CRC Biomarkers.....	32
1.2.6.1 Stool Biomarkers of CRC – DNA based.....	33
1.2.6.2 Stool Biomarkers of CRC – Non DNA based.....	37
1.2.6.3 Current Serum Biomarkers of CRC.....	38
1.3 Role of Proteomics in CRC Biomarker Discovery.....	43
1.3.1 Introduction to Proteomics.....	43

1.3.2 Proteomics Analysis.....	45
1.3.2.1 Introduction to Quantitative Proteomics.....	45
1.3.2.2 Separation step of Quantitative Proteomics.....	46
1.3.2.3.1 Top down Mass Spectrometry.....	48
1.3.2.3.2 Bottom up Mass Spectrometry/Shotgun Proteomics.....	49
1.4 Sample collection and storage in Quantitative Proteomics.....	52
1.5 Quantitative Proteomics – Separation techniques.....	53
1.5.1 Gel based.....	53
1.5.2 Non gel based Separation techniques.....	55
1.5.2.1 Capillary Electrophoresis (CE).....	55
1.5.2.2 Liquid Chromatography (LC).....	56
1.6 Protein Identification in Quantitative Proteomics – Mass Spectrometry.....	60
1.6.1 Introduction to mass spectrometry.....	60
1.6.2 Ion source.....	61
1.6.2.1 ESI.....	62
1.6.2.2 MALDI.....	65
1.6.3 Mass Analyser.....	68
1.6.3.1 Time of flight (TOF).....	69
1.6.3.2 Quadrapole.....	71
1.6.3.3 Ion Trap.....	72
1.6.3.3.1 Quadrapole Ion Trap.....	72
1.6.3.3.2 Kingdon Ion Trap.....	74
1.6.3.3.3 Fourier Transform Ion Cyclotron (FTIC).....	77
1.6.4 Detector.....	78
1.6.5 Tandem MS/MS.....	78
1.6.5.1 CID.....	81
1.6.5.3 PQD.....	83
1.6.5.4 ETD.....	83
1.6.5.5 ECD.....	84

1.7 Quantitative Proteomics.....	84
1.7.1 Metabolic Labelling Strategies.....	85
1.7.1.1 SILAC.....	85
1.7.2 Chemical Labelling Strategies.....	85
1.7.2.1 ICAT.....	85
1.7.2.2 ITRAQ.....	86
1.7.3 Label Free.....	88
1.8 Fractionation Technologies.....	91
1.8.1 Immunoaffinity Depletion.....	92
1.8.2 Multi-dimensional Separation – Strong Cation Exchange and Reverse Phase Chromatography.....	93
1.9 Bioinformatics – Processing and Analysis of Proteomic Data for Biomarker Discovery.....	94
1.9.1 Factors that affect Proteomic Profiling.....	95
1.9.2 Overview of Protein Identification.....	95
1.9.3 Data Analysis Methods.....	96
1.9.3.1 Database searching – MASCOT.....	97
1.9.3.1.1 Ion scoring systems.....	98
1.9.3.1.1.1 MOWSE.....	98
1.9.3.1.1.2 Probability based scoring.....	99
1.9.3.1.2 MASCOT search parameters.....	101
1.9.3.1.3 Validation of MS/MS searches in MASCOT.....	102
1.9.3.1.3.1 Decoy database and estimation of the false discovery rate (FDR)	102
1.9.3.1.3.2 Receiver operating curve (ROC).....	103
1.9.3.2 Data Analysis for Quantitative Proteomics.....	105
2.0 Post Translational Modifications – Glycosylation.....	107
2.0.1 Overview of Glycosylation.....	107
2.0.2 Glycosylation and Cancer.....	108
2.0.3 Glycomic Analysis of Glycans, Glycopeptides and Glycoproteins.....	110

2.0.3.1 Glycan Analysis.....	111
2.0.3.2 Glycoprotein and Glycopeptide Analysis.....	111
AIMS	114
MATERIALS AND METHODS	
2.1 Methods Employed for Whole Serum Analysis Utilising ITRAQ and Label Free Platforms.....	115
2.1.1 Serum collection.....	115
2.1.2 Depletion of Highly Abundant Serum Proteins.....	120
2.1.2.1 MARS Hu6, Agilent antibody depletion spin column.....	120
2.1.2.2 MARS Hu6, Agilent antibody depletion using 20 Mm TEAB as the depletion buffer.....	121
2.1.2.3 MARS Hu14, Agilent antibody depletion spin cartridge.....	121
2.1.2.4 Genway IgY14, LC10 and Supermix LC2 antibody depletion columns.....	122
2.1.3 Protein Precipitation/Desalting Methods.....	123
2.1.3.1 Acetone Precipitation.....	123
2.1.3.2 Buffer exchange with Agilent 5KDa molecular cut off weight markers.....	124
2.1.4 Protein Reduction.....	124
2.1.5 ITRAQ Labelling.....	125
2.1.6 Peptide Separation Methods.....	126
2.1.6.1 C18 Zip tips.....	126
2.1.6.2 Off-line SCX: Thermo Biobasic SCX column and Polysulfoethyl A SCX Column.....	127
2.1.6.3 Phenomenex Gemini 3umC18 150 x 2.0 mm reverse phase column.....	129
2.1.7 Serum preparation for label free analysis.....	130
2.1.8 Second dimension: Liquid chromatography-electrospray ionisation tandem mass spectrometry analysis of ITRAQ samples.....	131
2.1.8.1 LC Packing's nanoflow HPLC system interfaced with an applied biosystems QSTAR XL QqTOF mass spectrometer.....	131
2.1.8.2 Dionex UV3000 HPLC coupled with an LTQ Orbitrap XL using higher collisional dissociation (HCD) fragmentation.....	133

2.1.8.3 Dionex UV3000 HPLC coupled with an LTQ Orbitrap XL using Pulsed Q collision induced dissociation (PQD).....	134
2.1.9 Second dimension: Liquid chromatography-electrospray ionisation tandem mass spectrometry label free analysis.....	135
2.1.10 Quantitation and Identification of mass spectrometry derived ITRAQ and label free data.....	136
2.1.10.1 ITRAQ quantitation using MASCOT.....	136
2.1.10.2 Label free quantitation utilising Progenesis LC-MS.....	139
2.2 SDS-PAGE Analysis of Serum Profile.....	143
2.2.1 ID gel preparation.....	143
2.2.2 Gel band clean up.....	144
2.2.3 In-gel reduction/Alkylation.....	144
2.2.4 Tryptic digestion of gel band slices.....	144
2.2.5 Gel band extraction.....	145
2.2.6 LC-MS/MS analysis of gel pieces.....	145
2.2.7 Fingerprint identification of gel slice extractions.....	146
2.3 Sub-proteome Analysis – Glycosylation Analysis using Multi-lectin Affinity Chromatography.....	147
2.3.1 Preparation of buffers and multi-lectin affinity PD-10 disposable column.....	147
2.3.2 Elution of captured proteins.....	148
2.3.3 Concentration, reduction, alkylation and digestion of the generated glycoproteins.....	149
2.3.4 LC-MS/MS analysis of glycoproteins isolated using the multi-lectin affinity column.....	149
2.3.5 Fingerprint identification of glycoproteins following lectin immunoaffinity analysis.....	150

RESULTS

3.1 Comparison of Serum Depletion Methods.....	151
3.2 Comparison of Protein Precipitation/Desalting Methods.....	156
3.3 Comparison of Peptide Separation Methods.....	157
3.4 Optimisation of HPLC Gradient.....	162

3.4.1 HPLC gradient length optimisation.....	162
3.5 Optimisation of QSTAR Parameters.....	162
3.5.1 Effect of enhanced product ion scan duration on subsequent Protein identification.....	162
3.6 Optimisation of Orbitrap Settings.....	164
3.6.1 Optimal HCD collision energy for generation of ITRAQ reporter ions/fragmentation data.....	164
3.6.2 Comparison of QSTAR and Orbitrap for ITRAQ quantitation.....	167
3.7 Assessment of ITRAQ Labelling Error.....	168
3.8 Final ITRAQ Experiments.....	169
3.8.1 CRC pool results.....	169
3.8.2 Dukes A-C ITRAQ results.....	172
3.9 Results for CRC Pool Label Free Analysis.....	187
4.0 Results for Dukes Stage Label Free Analysis.....	196
4.1 Ingenuity Pathway Analysis (IPA).....	207
4.1.1 Core analysis.....	207
4.1.1.1 IPA core analysis for CRC proteins identified with ITRAQ quantitation.....	209
4.1.1.2 IPA core analysis for CRC label free quantitation.....	215
4.1.1.3 IPA core analysis for Dukes proteins identified with ITRAQ quantitation.....	221
4.1.1.3.1 Dukes A ITRAQ IPA analysis.....	221
4.1.1.3.2 Dukes B ITRAQ IPA analysis.....	231
4.1.1.3.3 Dukes C ITRAQ IPA analysis.....	243
4.1.1.4 IPA for Dukes proteins identified with label free quantitation.....	253
4.1.2 IPA biomarker analysis.....	263
4.1.2.1 Dukes A biomarker analysis (see Table 53).....	263
4.1.2.2 Dukes B biomarker analysis (see Table 54).....	264
4.1.2.3 Dukes C biomarker analysis (see Table 55).....	264

4.1.2.4 Combined Dukes A-C biomarker comparison.....	265
4.2 ID PAGE Analysis of Control and CRC Depleted Serum.....	276
4.3 Comparison of ITRAQ Label Free Quantitation and ID Gel Analysis.....	283
4.4 Results for Serum Glycosylation Analysis using Lectin Affinity Chromatography.....	285
DISCUSSION	
5.1 Overview of Proteomics.....	293
5.2 Study Population.....	296
5.3 Method Development of Quantitative Proteomic Platform.....	297
5.3.1 Quantitative Proteomics.....	300
5.3.1.1 ITRAQ.....	301
5.3.1.2 Label free.....	303
5.3.2 Depletion methods.....	306
5.3.3 Peptide separation methods.....	309
5.3.4 Optimisation of mass spectrometry parameters.....	310
5.3.5 ITRAQ labelling error.....	313
5.4 CRC Biomarkers.....	315
5.4.1 CRC ITRAQ and label free biomarkers.....	315
5.5 Dukes Biomarkers.....	324
5.5.1 Dukes ITRAQ and label free biomarkers.....	324
5.6 Gel.....	327
5.7 Sub-Proteome Analysis – Glycosylation.....	330
CONCLUSIONS and FUTURE WORK.....	333
APPENDIX.....	337
REFERENCES.....	395

