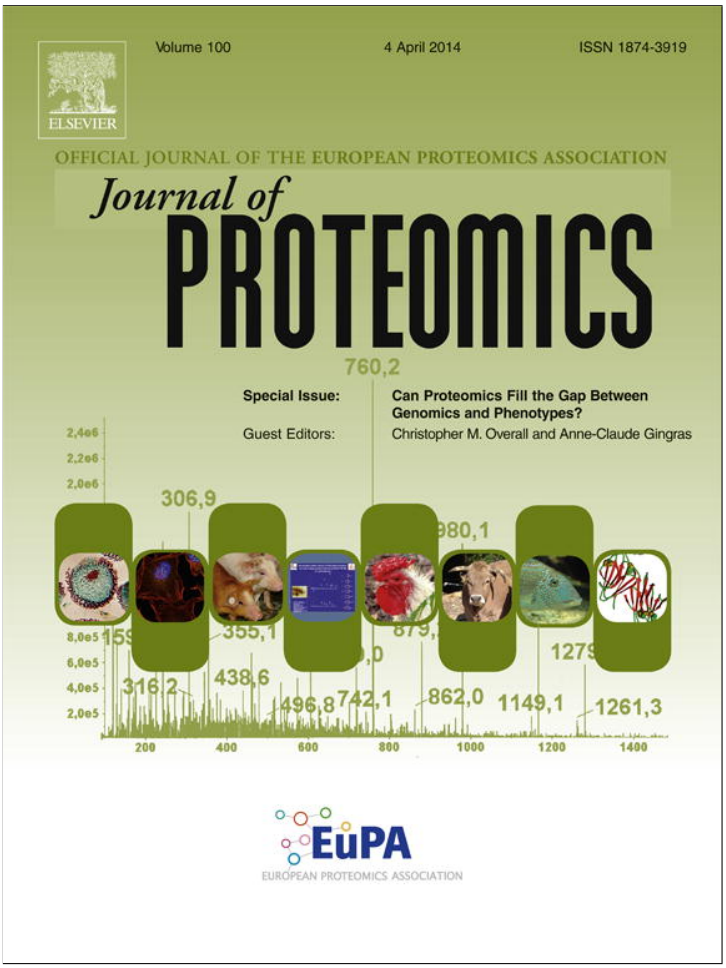

JOURNAL OF PROTEOMICS 100 (2014) 60–67

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Review

The Human Proteome Organization Chromosome 6 Consortium: Integrating chromosome-centric and biology/disease driven strategies☆

C.H. Borchers[a], J. Kast[b], L.J. Foster[c], K.W.M. Siu[d], C.M. Overall[e], T.A. Binkowski[f], W.H. Hildebrand[g], A. Scherer[h], M. Mansoor[i], P.A. Keown[i,j,*], for the Human Proteome Organization Chromosome 6 Consortium

[a]University of Victoria/Genome BC Proteomics Centre, Victoria, BC, Canada
[b]Biomedical Research Centre, University of British Columbia, Vancouver, BC, Canada
[c]Centre for High Throughput Biology, University of British Columbia, BC, Canada
[d]Centre for Research in Mass Spectrometry, York University, Ontario, Canada
[e]Centre for Blood Research, Faculty of Dentistry, University of British Columbia, Canada
[f]Midwest Centre for Structural Genomics, Argonne National Laboratory and Computation Institute, University of Chicago, USA
[g]Department of Microbiology and Immunology, University of Oklahoma, OK, USA
[h]Australian Genome Research Facility, Walter and Eliza Hall Institute, Parkville, Australia
[i]Department Medicine, University of British Columbia, Vancouver, BC, Canada
[j]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

ARTICLE INFO

Available online 8 August 2013

Keywords:
Human Proteome Project
Proteomics
Chromosome 6

ABSTRACT

The Human Proteome Project (HPP) is designed to generate a comprehensive map of the protein-based molecular architecture of the human body, to provide a resource to help elucidate biological and molecular function, and to advance diagnosis and treatment of diseases. Within this framework, the chromosome-based HPP (C-HPP) has allocated responsibility for mapping individual chromosomes by country or region, while the biology/disease HPP (B/D-HPP) coordinates these teams in cross-functional disease-based groups. Chromosome 6 (Ch6) provides an excellent model for integration of these two tasks. This metacentric chromosome has a complement of 1002–1034 genes that code for known, novel or putative proteins. Ch6 is functionally associated with more than 120 major human diseases, many with high population prevalence, devastating clinical impact and profound societal consequences. The unique combination of genomic, proteomic, metabolomic, phenomic and health services data being drawn together within the Ch6 program has enormous potential to advance personalized medicine by promoting robust biomarkers, subunit vaccines and new drug targets. The strong liaison between the clinical and laboratory teams, and the structured framework for technology transfer and health policy decisions within Canada will increase the speed and efficacy of this transition, and the value of this translational research.

☆ This article is part of a Special Issue: Can Proteomics Fill the Gap Between Genomics and Phenotypes?

* *Corresponding author at: Department of Pathology and Laboratory Medicine, University of British Columbia, Room 1559, Immunology, Vancouver General Hospital, Vancouver V5Z1M9, Canada.* Tel.: +1 604 875 4393; fax: +1 604 875 4911.
E-mail address: paul_keown@yahoo.com (P.A. Keown).

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2013.08.001

Biological significance
Canada has been selected to play a leading role in the international Human Proteome Project, the global counterpart of the Human Genome Project designed to understand the structure and function of the human proteome in health and disease. Canada will lead an international team focusing on chromosome 6, which is functionally associated with more than 120 major human diseases, including immune and inflammatory disorders affecting the brain, skeletal system, heart and blood vessels, lungs, kidney, liver, gastrointestinal tract and endocrine system. Many of these chronic and persistent diseases have a high population prevalence, devastating clinical impact and profound societal consequences. As a result, they impose a multi-billion dollar economic burden on Canada and on all advanced societies through direct costs of patient care, the loss of health and productivity, and extensive caregiver burden. There is no definitive treatment at the present time for any of these disorders.
The manuscript outlines the research which will involve a systematic assessment of all chromosome 6 genes, development of a knowledge base, and development of assays and reagents for all chromosome 6 proteins. We feel that the informatic infrastructure and MRM assays developed will place the chromosome 6 consortium in an excellent position to be a leading player in this major international research initiative.
This article is part of a Special Issue: Can Proteomics Fill the Gap Between Genomics and Phenotypes?

Contents

1. Introduction

Chromosome 6 (Ch6), a metacentric chromosome 171.11 Mbs in length, contains approximately 6% of the human genome [1]. The first gene map was completed in 2003, and current sequence data identify a total complement of between 2344 and 2780 genes, with an average density of 16.2 genes per Mb [2]. Between 1002 and1034 of these genes code for known, novel or putative proteins, and about 2.2% of the chromosome is occupied by exons with a mean length of 281 Bps. More than 350 other genes code for miRNA, snRNA, snoRNA and miscellaneous transcripts, while a further 700 are processed or unprocessed pseudogenes [2]. Recent studies have identified genes related to critical biological functions throughout the length of Ch6, of which the largest is the PARK2 gene on the q arm (1.4 Mb, 12 exons) [3–5]. These genes code for approximately 3000 known protein transcripts expressed in extracellular, intracellular or membrane compartments, many are involved in immunity, inflammation, neuronal activities and other critical cellular activities, of which key examples are presented in Table 1.

Of the several discrete regions within the chromosome, one of the most prominent is the extended major histocompatibility complex (eMHC). This 7.6 Mb super-region is located on the short arm of Ch6 and extends telomerically from RPL12P1 to HIST1H2AA [2]. The five sub-regions of the eMHC contain 523 genes, of which approximately 260 (50%) are expressed [2,6]. The eMHC is the most gene-rich region of the human genome, with a density of over 68 total genes and 35 protein coding genes per Mb. Several functional gene clusters have been defined within this extended region (six clusters and six superclusters) of which the two largest and potentially overlapping are the histone and tRNA genes. Transcripts of both are highly required in biological regulation and may be under selection pressure to cluster in association with the MHC [6]. The Human Leukocyte Antigen (HLA) genes located within the eMHC at 6p21.3 are critically related to infection, immunity and inflammation. The more than 200 genes within this hypervariable cluster are divided into 3 regions designated as class I, II and III. The HLA genes are typically highly polymorphic and exhibit tight linkage disequilibrium. More than 8000 alleles have now been identified within the HLA genes, coding for an estimated 6800 proteins of which only 2% can be serologically distinguished by current antibody methods.

1.1. *Role in human diseases*

Chromosome 6 is functionally associated with more than 120 major human diseases, including cancer, heart disease,

infectious, immune and inflammatory disorders and mental illnesses [7]. Many of these chronic diseases, examples of which are shown in Table 2, have a high population prevalence, devastating clinical impact and profound societal consequences, and as a result impose multi-billion dollar economic burdens on all advanced societies [8,9]. There is no definitive treatment at the present time for virtually any of these disorders.

Common structural defects in Ch6 result in defined clinical syndromes. Interstitial deletions within the 6p22-24 segment are associated with orofacial clefting, short neck, clinodactyly and heart and brain defects [10], while terminal deletions result in corneal opacity, facial dysgenesis and deafness [11]. Deletions within 6q are associated with cardiac anomalies, facial dysmorphisms and mental retardation. Duplications in 6p may also be associated with growth retardation, facial anomalies and mental retardation, while 6q duplications are associated with microcephaly, mental retardation, facial anomalies, and palatal, eye and genital anomalies. Such associations emphasize the disease relevance of genes encoded on Ch6 and the need to fully understand their biological roles and tissue expression patterns.

Table 1 – Selected proteins encoded on Ch6 with clinical or biological importance.

A. Extracellular proteins and subgroups in immunity and inflammation
Tumor necrosis factor α
Lymphotoxins A&B
Lymphocyte antigens 6 complexes, loci G5C, G6C, G5B, G6D
Complement factors C2, C4a, C4b, complement factor B (member of alternative pathway)
Vanins 1,2,3
Interleukin 1,2,3
Serum response factor
Apolipoproteins A and M
Vascular endothelial growth factor α
Connective tissue growth factor
Serpin protein 6B
Endothelin 1
Collagens IX-alpha1, X-alpha1, XI-alpha2, XII-alpha1, XIX-alpha1, XXI-alpha
Laminins lam-alpha2 & lam-alpha4
B. Membrane proteins
Opioid receptor,mu 1
GABA receptors 1, GABA receptor rho1, GABA receptor rho2
Serotonin receptors 1B & 1E
Glutamate receptors ionotropic kainate2, metabotropic1, metabotropic4
Interleukin receptors IL-20receptoralpha, IL-22 receptor,alpha2
Interferon gamma receptor
G-protein coupled receptors GPCR, GPCR family C, group 6,membrane A
C. Transcription factors and other proteins
3 PHD finger proteins members 1,3,10
Fyn and Fyn-related kinase
Ezrin
Flotillin 1
Gap junction proteins alpha 1, alpha 10, beta 7 and epsilon one
Natural cytotoxicity triggering receptors members 2&3
BCL-associated protein
Parkin 2
Prolactin
Vasoactive intestinal peptide(VIP)
Glycoprotein hormones, alpha polypeptide
Insulin-like growth factor 2 (somatomedine A) receptor
Hypocretin (orexin) receptor 2
Pepsinogen

Immune, inflammatory and degenerative diseases associated with specific sequence variations or multiple allelic heterogeneity include type I diabetes mellitus, multiple sclerosis, Alzheimer's disease, Parkinson disease, schizophrenia, rheumatoid arthritis and many other inflammatory or non-inflammatory disorders. The societal impact of these disorders is enormous (Table 3). The population prevalence of juvenile diabetes is estimated to be 1%; life expectance is reduced by 15 years in this disease; chronic ocular, renal and neural complications are typical, and the economic burden is around $2 billion annually in Canada [12]. The prevalence of multiple sclerosis in Canada is among the highest in the world, with societal economic costs of $1 billion per year [13]. Alzheimer's disease, schizophrenia, rheumatoid arthritis, celiac disease and others add to this burden of disorders associated with Ch6, with annual costs in the billions of dollars [14,15].

Other disorders may reflect altered expression of protein coding genes of Ch6. Increased expression of vascular endothelial growth factor (VEGF) occurs in POEMS syndrome (Crow–Fukase syndrome) [15,16] and VEGF is highly associated with metastatic cancers. PPIL1 gene expression is increased in colon cancer cells and reduction in its expression may help to suppress the growth of these malignant cells. BMP6 gene over expression is associated with aggressiveness prostatic cancer and its potential role as a prognostic predictor, while expression of the K1FC1 gene in the extended class II sub region of MHC is reported to be related to brain metastasis from non-small cell lung cancer.

Malignancy may be associated with structural deletions (e.g. lymphoblastic leukemia) [17,18], loss of heterozygosity or copy number variation [19,20], or defined risk loci which may function as tumor-suppressor genes [21–23]. Loss of heterozygosity in MHC genes has been closely related to acute lymphoblastic leukemia (ALL) [18], while a tumor suppressor gene at 6q15-21 has an important role in acute lymphoma type adult T-cell leukemia (ATL) and in childhood ALL [24]. Deletions in 6q21 occur in 7% of B-cell chronic lymphocytic leukemias (B-CLL) [25] and of 6q27 in 21% of patients with CLL, while deletions in the long arm of chromosome 6 are related to diffuse B-cell lymphoma in testis [26]. Ch6 transcription factor genes have been implicated solid organ tumors. The mesodermal specific tumor suppressor gene TCF2 is silent in malignancies of the head and neck or lung cancers, while the DNA-binding transcriptional repressor PHD Finger Protein 1 (PHF1) is associated with endometrial stromal cancer in recombination with other genes from chromosomes 7 and 10. Enzyme coding genes may enhance malignant risk through their detoxification activity or their direct cell cycle regulation and tumor suppressing effects.

Ch6 genes such as C6orf173 (CENPW) may influence somatic growth, the response to different pathogens or drugs, or may be protective against other diseases such as the non-HLA gene NFKBIL1 in the MHC region and HLA-DRB1 which is strongly protective against type 1 diabetes mellitus regardless of its linkage disequilibrium with HLA-DRB1.

1.2. *Team and technologies*

The Ch6 Consortium has integrated the initiatives of both the C-HPP and B/D-HPP to complete the proteomic mapping and

Table 2 – Principal chronic diseases associated with Chromosome 6. Disorders indicated by * are associated with exceptional clinical and societal burden.

- Alzheimer's disease *
- Ankylosing spondylitis *
- Autism *
- Behcet's disease *
- Bipolar disorder *
- Celiac disease *
- CHAR syndrome
- Complement deficiency
- Crohn's disease *
- Diabetes mellitus type 1 *
- Ehlers-Danlos syndrome
- Epilepsy *
- Fanconi anemia
- Hashimoto's thyroiditis *
- Macular degeneration *
- Maple syrup urine disease
- Multiple sclerosis *
- Narcolepsy *
- Nephritis *
- Neuroblastoma *
- Parkinson disease *
- Pemphigus vulgaris *
- Polycystic kidney disease *
- Porphyria
- Primary ciliary dyskinesia
- Psoriasis *
- Retinitis pigmentosa
- Rheumatoid arthritis *
- Schizophrenia *
- Spinocerebellar ataxia
- Sudden infant death syndrome
- Systemic lupus erythematosus *
- Tourette syndrome
- Viral resistance and response *

to discern the biological role of the relevant proteins in human disease. The consortium has developed an international collaboration combining clinomics, genomics, proteomics, terminomics and metabolomics platforms, and has selected a sequence of discrete targets of critical importance in relation to diseases closely linked to this chromosome. Coordinated projects within the overall program include the immunopeptidome, histone-binding proteins, the N and C-terminomes of the Ch6 proteins and other relevant targets of primary clinical and biological importance. Strategic approaches combine 3D-computerized modeling of proteins and bound peptides, large-scale in-vitro expression using bioreactor technology, and localization and quantitative monitoring of target tissues in normal healthy subjects and precisely phenotyped primary disease populations using nanobore liquid chromatology, MRM/MS and other advanced proteomic methods.

The structure and organization is shown in Fig. 1. The steering committee draws together expertise in the fields of genomics, proteomics, molecular sciences, statistics and bioinformatics, and clinical medicine, to contribute complementary and synergistic expertise. The Consortium includes clinical scientists from infectious disease, multiple sclerosis, rheumatology, diabetes, and stem cell and organ transplantation who provide precisely phenotyped subjects with relevant diseases and supervise their biological study. Extensive collaboration with international teams and close integration with partner academic institutions and industry are designed to enable efficient distribution of responsibilities, effective use of cutting edge resources, and rapid scaling of basic and applied research.

1.3. *Strategic approach*

The Ch6 HPP program will proceed in discrete stages, with clinical samples drawn from extensive partner biolibraries in principal areas of infection, autoimmunity, and alloimmunity used to compare normal and affected tissues in precisely phenotyped subjects for each of the targeted disease states. The four stages of the Ch6 HPP program will involve the development of MRM assays for all Ch6 proteins, the development of a Ch6 knowledge base, the generation of Ch6 reagents and primary data collection from a range of tissues known to express different Ch6 proteins to determine their N and C termini. The four stages will be integrated throughout the program to maintain a focus on unknown proteins and protein isoforms.

MRM assays will be generated for each protein encoded on Ch6 starting with extant data in public mass spectrometry databases. In order to ensure confident detection and quantitation, a minimum of 3 peptides will be targeted for each protein [27,28]. This strategy, which constitutes the most sensitive protein detection technique currently in use, has been successfully implemented in several member labs with limits of detection routinely reported with attomole sensitivity [29,30]. Peptides that would allow detection of specific splice variants will be targeted, while avoiding those covering regions of known polymorphism. As N and C termini data are collected by terminal amino isotope labeling of substrates (TAILS) [48,49,51] analyses SRMs for termini reflecting altered biological function will be generated in order to quantify the activity of different proteins where this is altered by proteolytic processing [52,53]. Previously unknown or uncharacterized proteins whose expression can be confirmed in this way will then be targeted in a second phase using peptide antibodies. Such antibodies will

Table 3 – Presentation, prevalence, and socioeconomic impact of selected major diseases associated with Ch6.

Rheumatoid disease
Autoimmune disorder causing relapsing, progressing inflammatory joint disease with deformity and incapacitation. Population prevalence estimated at almost one quarter million. Management is improving with biological therapeutics, but restoration of specific self tolerance is the ultimate goal. Economic burden is extreme, exceeding $5 billion per year.

Juvenile diabetes
Autoimmune disorder causing impaired glucose metabolism, leading to progressive blindness, renal failure, vascular damage and amputations. Prevalence estimated at 1% of population. Life expectancy reduced by at least 15 years. Management is improving but remains inadequate, with no cure. Economic burden estimated at $2 billion per year

Schizophrenia
Progressive, relapsing and destructive mental disease with compelling immunogenetic predilection causing psychological and psychiatric deterioration, institutionalization, and incarceration. Management is inadequate, and there is no current cure. Economic burden estimated at over $2 billion per year.

Alzheimer's disease
Most important form of degenerative neurological disease responsible for two thirds of cases dementia. Population prevalence estimated at over half a million and increasing rapidly with an aging society. There is currently no effective management, and no cure. Societal and economic burden is profound, estimated at $15 billion per year.

Multiple sclerosis
Progressive demyelinating neuroimmune disease leading to paralysis and immobility. Population prevalence in Canada is among the highest in the world with 75,000 sufferers. Current management strategies are oriented to delay disease progression, but no cure is available. Societal and economic burden is profound, estimated at $1 billion per annum.

be used to rapidly characterize the expression of these proteins in various cell and tissue types using a combination of immunoenrichment followed by liquid chromatography-multiple reaction monitoring (LC-MRM) analysis [29,31] or MALDI-MS [32], as well as more conventional but high throughput approaches such as tissue microarray analysis.

The immunopeptidome will form a specific research target within this framework, reflecting the enormous allelic polymorphism in the MHC class I and II regions and the diversity of immune responses across the human population. The Ch6 HPP Consortium will take advantage of its unique teams and expertise to measure self and foreign peptides presented by MHC molecules in a set of key immune and inflammatory disorders using a structured approach (Fig. 2).

A highly-scalable methodology, implemented in a leadership-scale supercomputing environment at the Argonne Leadership Computing Facility, will be used to accurately model bound peptides in HLA molecules. The iterative algorithm invokes an initial homology template-based methodology [33] and escalates into more advanced, physics-based approximations of protein–protein interactions [34] and estimations of binding free energy [35]. This method has been rigorously benchmarked against experimentally determined binding data from the Immune Epitope Database (IEDB) repository [36] and outperforms reported structural based prediction methods.

HLA immunopeptides will then be identified by nanobore liquid chromatography (nanoLC) coupled to tandem mass spectrometry [37–40]. Thousands of immunopeptides that originated from membranal HLA (mHLA) and sHLA molecules have been identified from cells and blood [40]. mHLA from tissue/cell lysates and sHLA from blood will be isolated using commercially-available antibodies and peptides released by stripping with acetic acid [41], concentrated and prepared for nanoLC-tandem mass spectrometry using state-of-the-art instrumentation. As HLA peptides are non-tryptic their dissociation under MS/MS will be less predictable than that of tryptic peptide [39,40,42]. Peptides will be identified using a two-step operation using shotgun proteomics with instrument-specific algorithms and de novo sequencing. The identification of HLA-bound peptides will be facilitated by the high resolution and high mass-accuracy measurements and knowledge of HLA binding motifs [43,44].

Multiple reaction monitoring (MRM) assays will be developed for all candidate MS antigens, commencing with known methods as a starting point with data drawn from on-line data repositories such as the Global Proteome Machine, PRIDE and PeptideAtlas. Bioinformatics tools will be used to predict the tryptic peptides most likely to make good MRM candidates, which will then be synthesized these in stable-isotope-labeled form and their detection empirically optimized via infusion electrospray [45–47].

In addition to examining proteins encoded by Ch6, studies will map the position and nature of histone proteins that scaffold them. Using a modified ChIP-Seq (chromatin immunoprecipitation with massively DNA sequencing) approach, the research will cartograph at the mononucleosome level, the relative position of H2A, H2B, H3, and H4 and will examine histone variant abundance and distribution. Given that the presence of specific histone variants have been correlated with the level of expression of genes related to various diseases, this initial map will provide a starting point to establish linkages between histone protein contents and protein expression. In parallel, studies will map the different post-translational modifications, including lysine methylation and acetylation, serine phosphorylation and arginine methylation, along this chromosome. Finally, mass spectrometry approaches will be employed to map protein complexes co-localizing with specific marks. This project will provide the first protein–protein network specific to Ch6, and the first systematic atlas of histone post-translational modifications on one chromosome. Mapping of these marks will be instrumental for establishing linkages between a given mark and a specific disease, and given that these marks frequently work in combinatorial fashion, the development of new antibodies that recognize combinatorial marks will provide important working tools for the community of chromatin biologists and epigeneticists.

In addition, studies will define the N and C-termini of all Ch6 proteins and their post-translational modifications using TAILS in both the N and C-TAILS variants: N-TAILS for original mature protein N termini [48,49] and C-TAILS for the protein-C-termini [50]. A TAILS-based workflow for the sensitive and reliable detection of N termini in complex human tissues has recently been established using erythrocytes as technically challenging source material [54]. In these procedures we deploy novel polymers for proteomics that we developed. Thus, to analyze the N-terminome, dendritic aldehyde polymers remove tryptic and C-terminal semi-tryptic peptides leaving the unbound naturally occurring acetylated or cyclized protein termini and the labeled mature protein N-termini by MS/MS. These dendritic polymers are a novel class of exceedingly efficient, highly functional (~3200 functionalities/molecule), high molecular weight (MW) (565 kDa) water-soluble polymers. Transcript analyses will be used to select tissues richly expressing Ch6 proteins and TAILS performed. For the Ch6 project new catch and release polymers will be developed for the selective enrichment of N and C termini and other PTM modified peptides in Ch6 proteins.

Results from the MRM analyses coupled with antibody-based screens will be used in developing the knowledge base, the third pillar of the HPP. While this base will be Ch6 specific, it will be aligned with the global effort. An informatics and bioinformatics platform for planning and coordination, assessment and management of resources, communications, information gathering, distribution of work, quality assurance and control, annotation and curation, sorting and prioritizing, statistical analysis, data representation, interpretation, deployment and distribution, usage and usability, and training is now in preparation. This will ensure collaboration with neXtProt, the HuPO database for the Human Proteome Project (HPP), and will help to define the key features of this data repository. Integration of existing data repositories to generate complete mRNA and protein lists of all genes encoded by Ch6 is now underway, and results will be made available via the Chromosome6.ca data portal for this program. The research community will be encouraged to share information on Ch6 genes and proteins for insertion into the data portal.

The last goal of the Ch6 HPP program will be to generate reagents for Ch6 gene products where there are none available. The knowledge base will help the consortium reach out to the research community and encourage sharing of currently available Ch6 reagents such as anti-peptide antibodies. The team will then concentrate on the development of reagents for

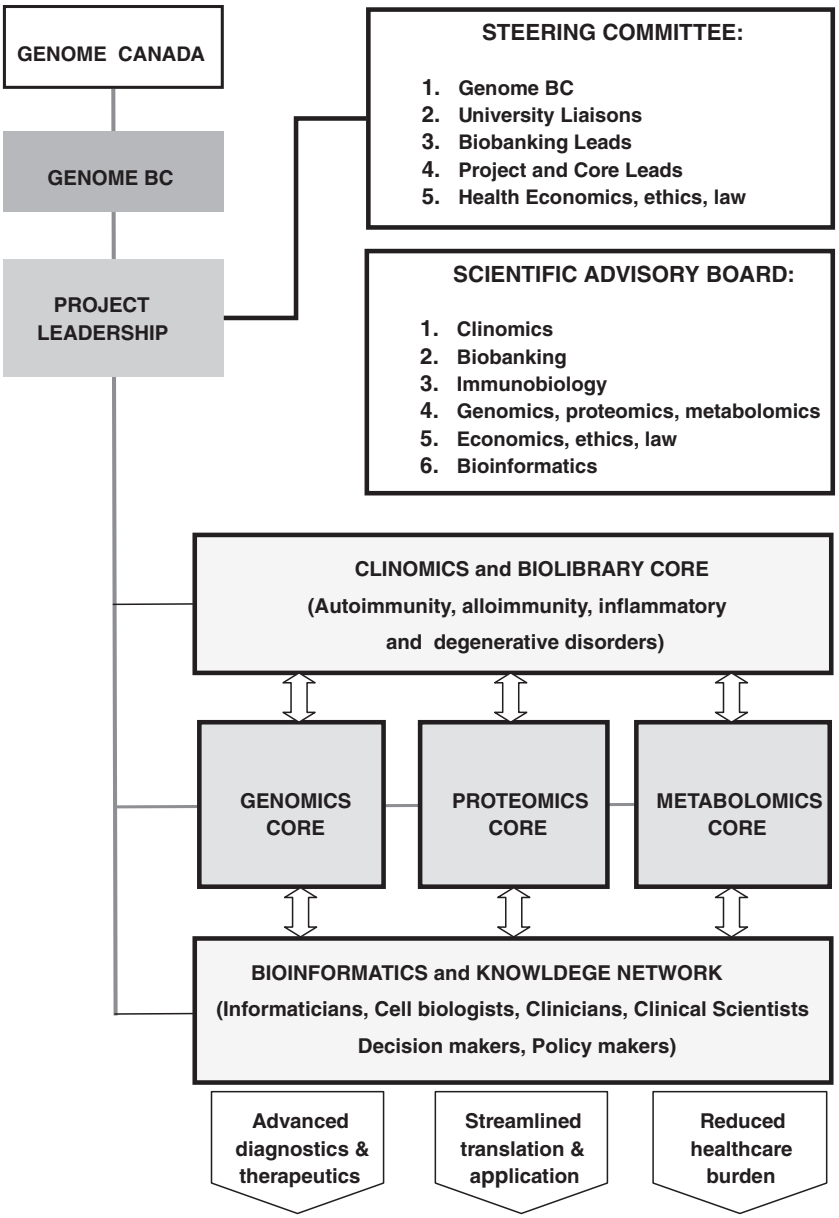

Fig. 1 – Structure and organization of the Chromosome 6 Consortium.

those Ch6 proteins which lack any means of protein detection. The aim is to have commercializable reagents for all Ch6 proteins which can then be used to define protein function and disease relevance.

1.4. *Expected impacts*

The benefits of the HPP will be far-reaching, providing the basis for personalized care, strategies for comprehensive disease management, and the opportunity to reduce the societal burden of human disease. Capitalizing on these opportunities, however, requires a carefully structured and rigorously coordinated approach to discovery and development.

Disorders associated with Ch 6 are among the most devastating, chronic and costly illnesses that affect society. Costs of care for these disorders are in the tens of billions of dollars annually in Canada alone, and the global impact is inestimable. These health costs are paralleled by catastrophic personal and productivity losses imposing an immense economic and physical burden on patients, families and society through destruction of personal health, disruption of family integrity, diversion of caregiver time and consumption of public health resources.

The unique combination of genomic, proteomic, metabolomic, phenomic and health services data now being drawn together within this program has enormous potential as we move towards personalized and evidence-based medicine and as the need for robust biomarkers, subunit vaccines and new drug targets grows increasingly acute. The combination of resources over the past two decades provides an exceptional opportunity to enable the development of precise and reliable biomarkers of inflammatory injury and novel targets

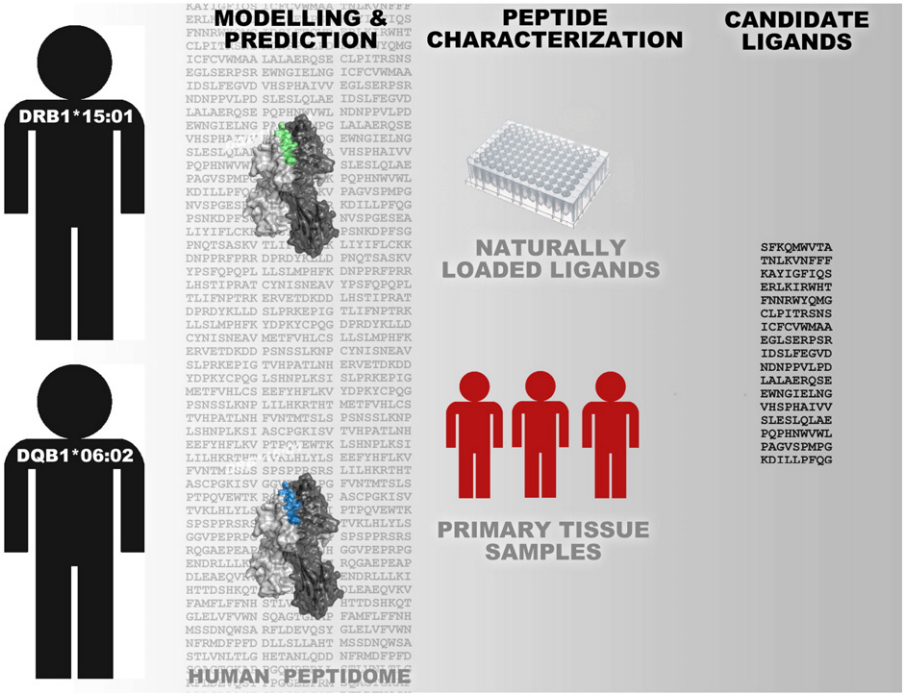

Fig. 2 – Modeling and characterization of the bound peptides in the human immunopeptidome.

for therapeutic development that will improve treatment and reduce tissue injury in the cardinal diseases, offering patients the potential to retain health and employment, contributing to human health and societal welfare. These innovations will be welcomed by clinicians world-wide, as demonstrated with our previous discoveries in different fields. The strong liaison between the clinical and laboratory teams, and the structured framework for technology transfer and health policy decisions within Canada, will increase the speed and efficacy of this transition, and the value of this translational research.

REFERENCES

[1] Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, et al. The DNA sequence and analysis of human chromosome 6. Nature 2003;425:805–11.

[2] Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2011. Nucleic Acids Res 2011;39:D800–6.

[3] Naj AC, Beecham GW, Martin ER, Gallins PJ, Powell EH, Konidari I, et al. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. PLoS Genet 2010;6:e1001130. http://dx.doi.org/10.1371/journal.pgen.1001130.

[4] Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. N Engl J Med 2009;360:1646–54.

[5] Veeriah S, Taylor BS, Meng S, Fang F, Yilmaz E, Vivanco I, et al. Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. Nat Genet 2010;42:77–82.

[6] Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet 2004;5:889–99.

[7] Communications LHNCfB: conditions related to genes on chromosome 6. Lister Hill National Center for Biomedical Communications; 2012.

[8] Wu EQ, Birnbaum HG, Shi L, Ball DE, Kessler RC, Moulis M, et al. The economic burden of schizophrenia in the United States in 2002. J Clin Psychiatry 2005;66:1122–9.

[9] AARDA, NCAPG. The cost burden of autoimmune disease: the latest front in the war on healthcare spending. American Autoimmune Related Disease Association (AARDA) and the National Coalition of Autoimmune Patient Groups (NCAPG); 2000.

[10] Lin RJ, Cherry AM, Chen KC, Lyons M, Hoyme HE, Hudgins L. Terminal deletion of 6p results in a recognizable phenotype. Am J Med Genet A 2005;136:162–8.

[11] DeScipio C. The 6p subtelomere deletion syndrome. Am J Med Genet C Semin Med Genet 2007;145C:377–82.

[12] Canadian Diabetes Association. The prevalence and costs of diabetes. Canadian Diabetes Association; 2009.

[13] Poppe AY, Wolfson C, Zhu B. Prevalence of multiple sclerosis in Canada: a systematic review. Can J Neurol Sci 2008;35:593–601.

[14] Arthritis Alliance of Canada: the impact of arthritis in Canada, today and over the next 30 years. www.arthritisalliance.ca.

[15] Alzheimer's Society of Canada: rising tide: the impact of dementia on Canadian society. http://www.alzheimer.ca/docs/RisingTide; 2010.

[16] Watanabe O, Arimura K, Kitajima I, Osame M, Maruyama I. Greatly raised vascular endothelial growth factor (VEGF) in POEMS syndrome. Lancet 1996;347:702.

[17] Hatta Y, Yamada Y, Tomonaga M, Miyoshi I, Said JW, Koeffler HP. Detailed deletion mapping of the long arm of chromosome 6 in adult T-cell leukemia. Blood 1999;93:613–6.

[18] McEvoy CR, Morley AA, Firgaira FA. Evidence for whole chromosome 6 loss and duplication of the remaining chromosome in acute lymphoblastic leukemia. Genes Chromosomes Cancer 2003;37:321–5.

[19] La Starza R, Aventin A, Matteucci C, Crescenzi B, Romoli S, Testoni N, et al. Genomic gain at 6p21: a new cryptic molecular rearrangement in secondary myelodysplastic syndrome and acute myeloid leukemia. Leukemia 2006;20:958–64.

[20] Huang L, Yu D, Wu C, Zhai K, Jiang G, Cao G, et al. Copy number variation at 6q13 functions as a long-range regulator and is associated with pancreatic cancer risk. Carcinogenesis 2012;33:94–100.
[21] Cozen W, Li D, Best T, Van Den Berg DJ, Gourraud PA, Cortessis VK, et al. A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. Blood 2012;119:469–75.
[22] López-Nieva P, Vaquero C, Fernández-Navarro P, Gonzalez-Sanchez L, Villa-Morales M, Santos J, et al. EPHA7, a new target gene for 6q deletion in T-cell lymphoblastic lymphomas. Carcinogenesis 2012;33:452–8.
[23] Oricchio E, Nanjangud G, Wolfe AL, Schatz JH, Mavrakis KJ, Jiang M, et al. The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. Cell 2011;147:554–64.
[24] Takeuchi S, Koike M, Seriu T, Bartram CR, Schrappe M, Reiter A, et al. Frequent loss of heterozygosity on the long arm of chromosome 6: identification of two distinct regions of deletion in childhood acute lymphoblastic leukemia. Cancer Res 1998;58:2618–23.
[25] Stilgenbauer S, Bullinger L, Benner A, Wildenberger K, Bentz M, Dohner K, et al. Incidence and clinical significance of 6q deletions in B cell chronic lymphocytic leukemia. Leukemia 1999;13:1331–4.
[26] Bosga-Bouwer AG, Kok K, Booman M, Boven L, van der Vlies P, van den Berg A, et al. Array comparative genomic hybridization reveals a very high frequency of deletions of the long arm of chromosome 6 in testicular lymphoma. Genes Chromosomes Cancer 2006;45:976–81.
[27] Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol 2009;27:633–41.
[28] Keshishian H, Addona T, Burgess M, Mani DR, Shi X, Kuhn E, et al. Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. Mol Cell Proteomics 2009;8:2339–49.
[29] DeSouza LV, Taylor AM, Li W, Minkoff MS, Romaschin AD, Colgan TJ, et al. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. J Proteome Res 2008;7:3525–34.
[30] Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, et al. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics 2009;8:1860–77.
[31] Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J Proteome Res 2004;3:235–44.
[32] Reid JD, Holmes DT, Mason DR, Shah B, Borchers CH. Towards the development of an immuno MALDI (iMALDI) mass spectrometry assay for the diagnosis of hypertension. J Am Soc Mass Spectrom 2010;21:1680–6.
[33] Krivov GG, Shapovalov MV, Dunbrack J, Roland L. Improved prediction of protein side-chain conformations with SCWRL4. Proteins 2009;77:778–95.
[34] Luo W, Pei J, Zhu Y. A fast protein–ligand docking algorithm based on hydrogen bond matching and surface shape complementarity. J Mol Model 16: 903–913.
[35] Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–97.
[36] Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, et al. A community resource benchmarking predictions of peptide binding to MHC-I molecules. PLoS Comput Biol 2006;2:e65.
[37] Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, et al. Pillars article: characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 1992. 255: 1261–1263. J Immunol 2007;179:2669–71.
[38] Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, et al. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A 2006;103:14889–94.
[39] Hawkins OE, Vangundy RS, Eckerd AM, Bardet W, Buchli R, Weidanz JA, et al. Identification of breast cancer peptide epitopes presented by HLA-A*0201. J Proteome Res 2008;7:1445–57.
[40] Bassani-Sternberg M, Barnea E, Beer I, Avivi I, Katz T, Admon A. Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc Natl Acad Sci U S A 2010;107:18769–76.
[41] Yu H, Karunakaran KP, Kelly I, Shen C, Jiang X, Foster LJ, et al. Immunization with live and dead *Chlamydia muridarum* induces different levels of protective immunity in a murine genital tract model: correlation with MHC class II peptide presentation and multifunctional Th1 cells. J Immunol 2011;186:3615–21.
[42] Aebersold R, Goodlett DR. Mass spectrometry in proteomics. Chem Rev 2001;101:269–95.
[43] Shastri N, Schwab S, Serwold T. Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules. Annu Rev Immunol 2002;20:463–93.
[44] Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 1991;351:290–6.
[45] Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GV, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics 2012;12:1222–43.
[46] Domanski D, Smith DS, Miller CA, Yang Y, Jackson AM, Cohen-Freue G, et al. High-flow multiplexed MRM-based analysis of proteins in human plasma without depletion or enrichment. Clin Lab Med 2011;31:371–84.
[47] Percy AJ, Chambers AG, Yang J, Domanski D, Borchers CH. Comparison of standard-flow and nano-flow liquid chromatography systems for MRM-based quantitation of putative plasma biomarker proteins. Anal Bioanal Chem 2012;404:1089–101.
[48] Kleifeld O, Doucet A, auf dem Keller U, Prudova A, Schilling O, Kainthan RK, et al. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat Biotechnol 2010;28:281–8.
[49] Kleifeld O, Doucet A, Prudova A, auf dem Keller U, Gioia M, Kizhakkedathu JN, et al. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nat Protoc 2011;6:1578–611.
[50] Schilling O, Barré O, Huesgen PF, Overall CM. Proteome-wide analysis of protein carboxy termini: C terminomics. Nat Methods 2010;7:508–11.
[51] Lange PF, Overall CM. TopFIND, a knowledgebase linking protein termini with function. Nat Meth 2011;8:703–4.
[52] Huesgen PF, Lange PF, Ovarall CM. Ensembles of protein termini are sensitive and specific proteolytic signature biomarkers of disease. J Prot 2013 [in press].
[53] Fahlman RP, Chen W, Overall CM. Absolute proteomic quantification of the activity state of proteases and proteolytic cleavages using proteolytic signature peptides and isobaric tags. J Prot 2014;100:79–91.
[54] Lange PF, Huesgen PF, Overall CM. Annotating N termini for the Human Proteome Project: N termini differentiate stable processed protein species from degradation remnants in the human erythrocyte proteome. J Prot 2013 [in press].