Characterization of patients with combined methylmalonic aciduria and homocystinuria diagnosed as *cblC*

Alison Brebner

Department of Human Genetics
McGill University
Montreal, Quebec, Canada

August 15, 2014

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Masters of Science

Abstract

Vitamin B_{12} (cobalamin) is essential for two mammalian enzymes: methylmalonylCoA mutase and methionine synthase. To date, there are nine described inborn errors of cobalamin metabolism which result in methylmalonic aciduria and homocystinuria, either alone or in combination. The most common, c*blC,* is caused by mutations in the *MMACHC* gene and results in the combined metabolic phenotype. Our laboratory has accumulated 339 *cblC* patient fibroblast samples; of these, 50 had one or no pathogenic identified mutations in *MMACHC*. DNA from these patients was subjected to Sanger sequencing of *MMACHC*. Four novel mutations, c.566G>A, c.158T>C, c.292C>T and c.551_554dupCTAC were found, along with an uncharacterized deletion of exon 2 and a presumed regulatory mutation causing allelic dropout of one allele. After this analysis, there still remained 13 samples with no mutations identified. Sanger sequencing identified the common mutation in *LMBRD1* in one patient, thus classifying it as *cblF*. Eleven had mutations in the transcriptional co-regulator *HCFC1* and one had a homozygous missense mutation in its binding partner *THAP11*. These patients were studied further as mutations in *HCFC1* (*cblX*) and *THAP11* resulted in a *cblC* phenocopy. Clinically they had a very severe neurological phenotype and milder biochemical findings than classical *cblC* patients. Whole transcriptome profiling using RNA sequencing was performed to compare *cblX* and the *THAP11* mutant fibroblasts against

control samples. Down-regulation of several same genes was noted in both *cblX* and *THAP11* cells, including *MMACHC* and *TMOD2*. *TMOD2* has been implicated in neuronal maintenance and differentiation. In conclusion, this thesis describes two novel genes causing inborn errors of cobalamin metabolism and four novel mutations in *MMACHC*.

Résumé

La vitamine B_{12}, également appelée cobalamine, est essentielle au fonctionnement de deux enzymes chez les mammifères: la methylmalonylCoA mutase et la méthionine synthase. À ce jour, neuf erreurs innées du métabolisme de la cobalamine ont été décrites entrainant soit une acidurie méthylmalonique soit une homocystinurie, soit les deux combinées. L'erreur innée la plus commune du métabolisme de la cobalamine, c*blC,* est causée par des mutations du gène *MMACHC* et aboutit à la manifestation d'un phénotype métabolique combiné. Notre laboratoire a accumulé 339 échantillons de fibroblastes issus de patients souffrant d'un déficit de *cblC*; parmi eux, 50 possédaient soit une seule des mutations pathogéniques identifiées du gène *MMACHC* soit aucune. L'ADN de ces patients a été soumis au séquençage du gène *MMACHC* par la méthode de Sanger. De nouvelles mutations ont été trouvées dont quatre aboutissent à une c.566G>A, c.158T>C, c.292C>T ou à une c.551_554dupCTAC ainsi qu'une nouvelle délétion non caractérisée au niveau de l'exon 2, et une mutation présumée avoir une fonction régulatrice et qui provoque une perte allélique. Finalement, 13 échantillons ne présentent pas mutations de *MMACHC*. Parmi eux, un appartenait à un patient ayant un déficit en cobalamine de type F (*cblF*) et une mutation commune dans le gène *LMBRD1* tel que révélé par le séquençage; onze avaient des mutations dans le co-régulateur transcriptionnel *HCFC1*; et un autre présentait une mutation faux-sens du gène *THAP11*, partenaire de liaison de HCFC1. Les patients avec des

mutations des gènes *HCFC1* (*cblX*) et *THAP11* sont des phénocopies *cblC*, causé par dérèglement de la transcription de ces deux gènes.

L'objectif suivant était d'étudier des patients présentant un défaut de transcription causé par la mutation des gènes *HCFC1* et *THAP11*. Sur le plan clinique, ces patients souffraient d'un phénotype neurologique très sévère mais des résultats modérés aux tests biochimiques. Une étude transcriptomique par séquençage d'ARN a été effectuée afin de comparer des échantillons de fibroblastes avec mutations des gènes *HCFC1* (*cblX*) ou *THAP11* à ceux provenant des contrôles. L'analyse a montré que dans les cellules *cblX* et *THAP11*, l'expression d'un même groupe de gènes était diminuée incluant *MMACHC* et *TMOD2*, un gène impliqué dans le maintien et la différenciation neuronale. En conclusion, cette thèse montre pour la première fois que les mutations des gènes *HCFC1* et *THAP11* cause une maladie héréditaire du métabolisme de la cobalamine ainsi que quatre nouvelles mutations dans *MMACHC*.

Table of Contents

Abstract 2
Résumé 4
Table of Contents 6
List of Abbreviations 9
List of Figures 11
List of Tables 12
Acknowledgements 13
CHAPTER 1 14
Introduction 14
1.1 Vitamin B_{12} 14
1.1.1 Structure 14
1.1.2 Dietary Requirement of Cbl 15
1.2 Cobalamin Absorption 17
1.2.1 Non-Heritable Defects of Cobalamin Absorption 19
1.2.2 Heritable Defects of Cobalamin Absorption, Transport and Cellular Uptake 20
1.2.2.1 Imerslund-Gräsbeck syndrome 20
1.2.2.2 Hereditary Intrinsic Factor Deficiency 21
1.2.2.3 Transcobalamin Deficiency 21
1.2.2.4 Haptocorrin Deficiency 22
1.2.2.5 Transcobalamin Receptor Deficiency 22
1.3 Intracellular Cobalamin Metabolism 23
1.3.1 Inborn Errors of Cellular Cobalamin Metabolism 24
1.3.1.1 Cellular Diagnosis 25
1.3.1.2 Somatic Cell Complementation Analysis 26
1.3.1.3 Methylmalonic Aciduria 29
1.3.1.4 Homocystinuria 32
1.3.1.5 Combined Methylmalonic Aciduria and Homocystinuria 34
1.4 The *cblC* Inborn Error of Cobalamin Metabolism 36
1.4.2 Clinical Symptoms 37
1.4.3 Outcome and Treatment 38
1.4.4 *MMACHC* Gene 40
1.4.4.1 *MMACHC* Genotypes in *cblC* Patients 40
1.4.4.2 MMACHC Function 41

1.5 HCFC1 and THAP11: Transcriptional regulators 41
1.5.1 Eukaryotic Transcriptional Regulation 42
1.5.2 Transcriptional Regulators and Mendelian Disease 43
1.5.3 HCFC1 46
1.5.4 THAP11 48
1.6 Rationale and Objectives of Study 50
CHAPTER 2 52
Materials and Methods 52
2.1 Patient cells used in this study 52
2.2 Cell Culture 52
2.3 DNA Extraction, Primer Design, PCR and Sequencing 53
2.4 RNA Sequencing 54
2.5 Restriction Digest Analysis of Common c.344C>T Mutation 55
2.6 Statistical Analyses 56
CHAPTER 3 57
Results 57
3.1 Sequencing of *MMACHC* in samples diagnosed as *cblC* 57
3.1.1 Mutation Analysis of Novel Mutations in *MMACHC* 59
3.1.1.1 Patients' Phenotype 59
3.1.2 Patients with no mutations in *MMACHC* 61
3.2 Characterization of patients with the *cblX* inborn error 62
3.2.1 Clinical Features of *cblX* Patients 63
3.2.2 Biochemical Features of *cblX* Patients 65
3.2.3 Haplotype Analysis of Patients with the Common c.344C>T Mutation in *HCFC1* 66
3.2.4 Carrier Testing of Mothers and Sisters of *cblX* Patients 67
3.2.5 Restriction Digest Analysis of the Common c.344C>T Mutation 68
3.3 Characterization of a Patient with Mutations in *THAP11* 70
3.3.1 Clinical and Biochemical Features of Patient WG2110 70
3.3.2 *THAP11* Mutation Analysis 71
3.4 Transcriptome Analysis of *cblX* and *THAP11* Patient Fibroblasts 73
CHAPTER 4 75
Discussion 75
4.1 Novel Mutations in *MMACHC* 75
4.2 *cblX* and *THAP11* Patients 78

4.3 Lack of Complementation Between *cblC* and *cblX* 80
4.4 RNA Sequencing Studies 83
References 86
APPENDIX A 101
PCR Conditions and Primer Sequences 101
APPENDIX B 104
Publications, Presentations and Abstracts 104
APPENDIX C 107
Article Yu *et al.* 2013 107

List of Abbreviations

AdoCbl: 5’-deoxyadenosylcobalamin

ATP: adenosine triphosphate

ATR: cobalamin adenosyltransferase

ATR-X: X-linked α-thalassemia with mental retardation

Bp: base pair

C3: propionylcarnitine

Cbl: cobalamin, vitamin B_{12}

CBS : cystathionine beta-synthase

ChIP seq: chromatin immunoprecipitation sequencing

CMAMMA: combined malonic aciduria and methylmalonic aciduria

CNCbl: cyanocobalamin

DMB: 5,6-dimethylbenzimidazole

ES cells: embryonic stem cells

FTT: failure to thrive

GFP: green fluorescent protein

HCFC1: host cell factor 1, HCF-1

HPLC: high performance liquid chromatography

IF: intrinsic factor

IM: intramuscular

IUGR: intrauterine growth retardation

Kb: kilobase

OHCbl: hydroxocobalamin

MAF: minor allele frequency

MCM: methylmalonyl CoA mutase

MeCbl: methylcobalamin

MeTHF: 5-methyltetrahydrofolate

MMA: methylmalonic acid

MMAuria: methylmalonic aciduria

MODY: maturity onset diabetes of the young

MS: methionine synthase

MSR: methionine synthase reductase

NBS: newborn screen

NMR: nuclear magnetic resonance

PCR: polymerase chain reaction

PEG: polyethylene glycol

RNA pol II: RNA polymerase II

RNA seq: RNA sequencing

SNP: single nucleotide polymorphism

TC: transcobalamin

TCblR: transcobalamin receptor

THAP : thanatos-associated protein

THF : tetrahydrofolate

UTR : untranslated region

VNTR : variable number tandem repeat

List of Figures

Figure 1: The Molecular Structure of Cobalamin 15

Figure 2: Dietary Cobalamin Absorption 18

Figure 3: Intracellular Cobalamin Metabolism 24

Figure 4: Somatic Cell Complementation Analysis 27

Figure 5: Biochemical Features of *cblX* versus *cblC* Patients 66

Figure 6: Restriction Digest Analysis of *HCFC1* c.344C>T 69

Figure 7: Multi-sequence alignment of human THAP proteins 71

Figure 8: Multi-species alignment of THAP11 72

Figure 9: Results of RNA sequencing experiments in control and *cblX* fibroblasts 74

List of Tables

Table 1: Summary of results from Sanger sequencing of the *MMACHC* gene in *cblC* patients……58

Table 2: Novel mutations in *MMACHC*……...59

Table 3: Patients with novel mutations in *MMACHC* – Age of Onset and Phenotype……61

Table 4: *cblX* Clinical findings and mutations in *HCFC1*……64

Table 5: Haplotype Analysis of patients with the common c.344C>T mutation in *HCFC1*……67

Acknowledgements

Firstly I'd like to thank my supervisor Dr. David Rosenblatt for having me in the lab these past three years. Your constant support, guidance and contagious enthusiasm made my Masters a productive and enjoyable experience.

I would also like to thank Dr. David Watkins. Your scientific advice, invaluable editing skills and entertaining stories will always be greatly appreciated.

Thank you to my supervisory committee, Dr. Tomi Pastinen and Dr. Pierre Moffatt for your help and guidance in my project. Your feedback and scientific insight were very helpful.

Thank you to Gail Dunbar for teaching me tissue culture and providing me with all the cell lines necessary to undertake this project. I would also like to thank Maria Galvez, Leah Ladores and Angeline Boulay for their help around the lab. Thank you to the amazing Human Genetics staff, Ross, Laura and Kandace. You made any administrative task simple and stress free.

This work would not have been possible without our collaborators. For the *cblX* project, from the University of Colorado: Dr. Tamim Shaikh and Hung-Chun Yu, from the National Institutes of Health: Dr. Charles Venditti, Dr. Jennifer Sloan and Nate Achilly and from Children's Hospital Zürich: Dr. Brian Fowler and Dr. Matthias Baumgartner. For the transcriptome studies: Dr. Tomi Pastinen, Shu-Huang Chen, Dr. Véronique Adoue and Dr. Stephan Busche. Also to all the physicians who sent us patient samples over the years. Thank you for Timothy Johns for performing all cell immortalizations and Shamisa Honarmand for teaching me how to do retroviral transductions.

To all my lab mates that have made my time in the lab enjoyable, thank you. Especially to my other two muskateers, Mihaela Pupavac and Wayne Mah for our insightful discussions. To Tracy Wang and Francis Petrella for all the tip boxes you filled. And a special thank you to Peg Illson, Laura Dempsey-Nunez and Jae Kim who initially welcomed me into the lab and were always so supportive and helpful. To everyone else who I have spent time with over the years, Tania Cruz, Justin Deme, Maira Moreno, Lena Dolman and Moria Belanger. Thank you to Elie Akoury for translating my French abstract.

And finally, thank you to Suleman, Mom, Dad, Karolyn and Andrew.

CHAPTER 1
Introduction

1.1 Vitamin B_{12}

1.1.1 Structure

Vitamin B_{12} (cobalamin, Cbl) is an organometallic compound that is required for human metabolism but is only synthesized by bacteria and archaea[1]. The structure of Cbl was solved in 1955 by Dorothy Hodgkin (Figure 1)[2]. It consists of a central cobalt atom that can have up to six ligands bound to it: four are nitrogen atoms that are part of the planar corrin ring; the α-axial ligand is a nitrogen from 5,6-dimethylbenzimidazole (DMB) phosphoribosyl that is covalently bonded to the corrin ring; and the β-axial ligand (R-group) varies. In the base-on conformation the DMB group is bound to the central cobalt atom[3]. In the base-off conformation the DMB is displaced from the central cobalt atom[4]. Cbl has three oxidation states: fully oxidized Co^{+++} (cob(III)alamin), Co^{++} (cob(II)alamin) and the fully reduced Co^{+} (cob(I)alamin)[1]. The pharmaceutical form of Cbl, which does not occur in nature, is cyanocobalamin (CNCbl), which has a cyanide group as the R-group. The natural derivatives include hydroxocobalamin (OHCbl), methylcobalamin (MeCbl), 5'-deoxyadenosylcobalamin (AdoCbl) and glutathionylcobalamin[1].

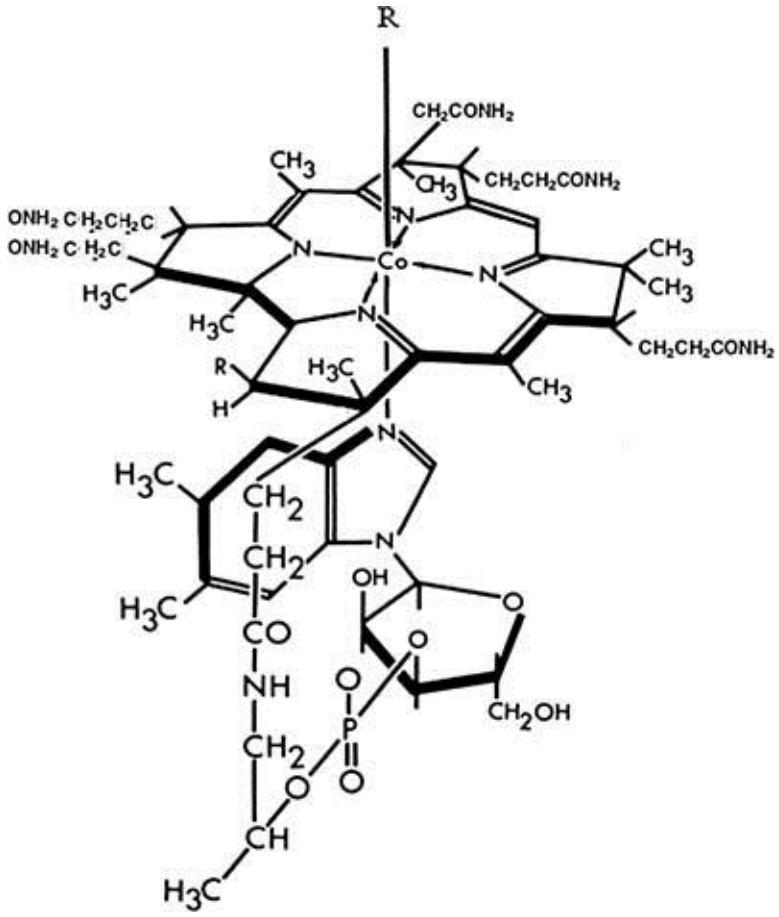

Figure 1: The Molecular Structure of Cobalamin

The central cobalt atom is bound to the planar corrin ring, the 5,6-dimethylbenzamidazole base in the lower axial position and the R group in the upper axial position. This structure is in the base-on conformation. Figure taken from Watkins and Rosenblatt, 2011[1].

1.1.2 Dietary Requirement of Cbl

Cbl is only synthesized by bacteria and archaea and must be obtained through the diet from animal products, as plants do not use or store Cbl. The recommended daily intake of Cbl for adults is 2.4µg per day[5]. Cbl is necessary for human metabolism. Its derivatives are required by two cellular enzymes: AdoCbl is necessary for the function of methylmalonyl CoA mutase (MCM) and MeCbl is required by methionine synthase (MS).

MCM converts methylmalonyl CoA, a product of the breakdown of branched-chain amino acids, odd chained fatty acids and cholesterol, into succinyl CoA, which can subsequently be used in the Krebs cycle. MS converts the amino acid homocysteine into methionine by methylation[6]. The Cbl metabolic pathway overlaps with that of folate. The enzyme MS simultaneously catalyzes the reaction converting methyltetrahydrofolate (MeTHF) to tetrahydrofolate (THF)[6].

Cbl deficiency is caused by multiple factors. Since this vitamin is found uniquely in animal products, a strict vegan diet without supplementation can result in a deficiency[7]. Defects in absorption that ultimately result in a Cbl deficiency such as pernicious anemia, decreased gastric acidity and gastric bypass surgery will be discussed in section 1.2.1.

The clinical findings of Cbl deficiency are quite severe. Cbl deficiency often results in a functional folate deficiency. Since the function of MS is blocked, the cofactor MeTHF cannot be metabolized. The reaction creating MeTHF through the enzyme MTHFR is irreversible under physiologic conditions, which results in the accumulation of this unavailable form of folate. This results in a “folate trap” and an absence of the folate forms that are necessary for DNA synthesis[8]. This process is very important in hematopoiesis, meaning a deficiency in Cbl often causes megaloblastic anemia. In this condition, red blood cells are macrocytic due to impaired production. The inhibition of DNA synthesis during the differentiation of the hematopoetic stem cells results in cell growth without

division, therefore large macrocytic cells are released rather than small, differentiated red blood cells[7].

In adults, the hematological findings are often accompanied by neurological abnormalities such as spinal cord or cerebral demyelination[7], which results in psychosis or memory loss[9]. These symptoms can be reversible upon supplementation with Cbl. If a mother has asymptomatic Cbl deficiency and is exclusively breast feeding, her infant can present with findings such as irritability, feeding difficulties, anemia and developmental delay. If diagnosis is too late, the developmental delay may be permanent[7].

1.2 Cobalamin Absorption

Cbl is absorbed in the intestine, transported in the blood, and metabolized in the cell. The absorption process can be seen in Figure 2. Most dietary Cbl is bound to protein. The protein-bound Cbl is ingested, and the acidity and proteolytic enzymes of the stomach cause the release of Cbl. It then binds haptocorrin, a glycoprotein present in saliva and gastric fluids[10]. Haptocorrin is digested in the duodenum by pancreatic proteases and Cbl subsequently binds intrinsic factor (IF), a glycoprotein secreted by the parietal cells of the stomach[11]. This IF-Cbl complex can bind the cubam receptor in the distal ileum. The cubam receptor is made of two proteins, cubilin and amnionless. The IF-Cbl complex is internalized by receptor mediated endocytosis at the apical surface of enterocytes[12]. Cbl exits enterocytes and subsequently enters the blood stream, where it

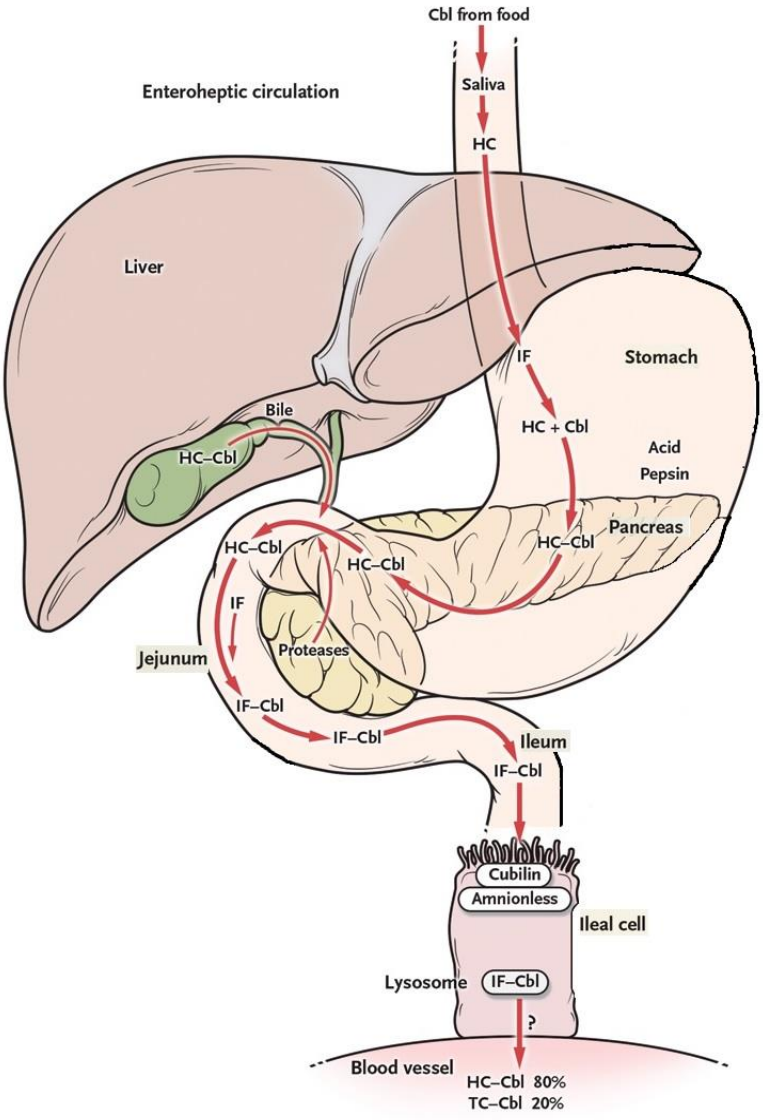

Figure 2: Dietary Cobalamin Absorption

Cbl from food enters bound to protein, which it dissociates from in the stomach and binds to haptocorrin (HC). Haptocorrin is digested by pancreatic proteases, the Cbl is released and binds intrinsic factor (IF). This complex binds the cubam receptor in the distal ileum and is released into the circulation where it remains bound to both HC and transcobalamin (TC). Figure modified from Stabler *et al.*[9]

circulates bound to transcobalamin (TC)[12]. TC-Cbl can bind the transcobalamin receptor (TCblR) on the cell and is engulfed by receptor-mediated endocytosis.

1.2.1 Non-Heritable Defects of Cobalamin Absorption

There are several defects of Cbl absorption that are non-heritable. These disruptions in absorption normally occur in the gastrointestinal tract. For example, disruption of the binding of Cbl to intrinsic factor can cause Cbl deficiency with a normal dietary intake. This can be caused by the auto-immune condition pernicious anemia. In this disease, antibodies are present against the gastric parietal cells or intrinsic factor causing their destruction[11]. Another cause of Cbl malabsorption due to its inability to bind IF, is quite common in the elderly. These patients do not have anti-IF antibodies, which rules out pernicious anemia as a cause. This malabsorption may be caused by atrophic gastritis, which results in decreased gastric acidity[7]. In normal individuals, the acidic environment of the stomach is necessary for the dissociation of Cbl from protein complexes found in food. The Cbl then normally binds IF for absorption. In patients with a less acidic gastric environment, there is inefficient absorption of Cbl in the ileum due to the decreased binding of Cbl to IF. Thirdly, Cbl malabsorption can be caused by gastrointestinal surgeries such as total gastrectomy, ileal resection or gastric bypasses. These surgeries can result in a loss of the parietal cells and/or loss of ileal cells that are responsible for uptake of the IF-Cbl complex. Some patients that

undergo these surgeries are prescribed proton-pump inhibitors, that result in a lower gastric acidity resulting in even more efficient absorption of Cbl[13].

1.2.2 Heritable Defects of Cobalamin Absorption, Transport and Cellular Uptake

There are several heritable defects that perturb either Cbl absorption, transport or cellular uptake. All the disorders characterized to date are very rare and exhibit an autosomal recessive mode of inheritance. These patients, with the exception of certain TCblR deficient patients, do not have a cellular biochemical phenotype when measuring the function of both Cbl-dependent enzymes. They do tend to have excretion of both metabolites, methylmalonic acid (MMA) and homocysteine. All the known conditions are described in more detail below.

1.2.2.1 Imerslund-Gräsbeck syndrome

Imerslund-Gräsbeck syndrome (OMIM #261100) results in Cbl malabsorption in the ileum. It is a pediatric disorder and patients generally present with megaloblastic anemia, neurological findings of variable severity, and often proteinuria. The protein excretion is unique among inborn errors of cobalamin and distinguishes this disorder from nutritional deficiencies or an absence of intrinsic factor. The patients' symptoms are reversible upon treatment with intramuscular (IM) Cbl injections, except for the proteinuria[14]. To date, patients have been described with mutations in

each of the two genes that make up the cubam receptor, *CUBN*[15] and *AMN*[16].

1.2.2.2 Hereditary Intrinsic Factor Deficiency

Hereditary IF deficiency (OMIM #261000) is caused by mutations in the *GIF* gene[17]. The disease presentation is different from patients with pernicious anemia as they have no antibodies to IF and their gastric mucosa is normal[18]. Clinically, patients typically present in childhood with megaloblastic anemia, however no proteinuria is present. This disorder is also treated with parenteral Cbl supplementation, bypassing the absorption in the ileum. Prompt treatment usually results in a reversal of any clinical symptoms of Cbl deficiency the patient may have[17].

1.2.2.3 Transcobalamin Deficiency

Transcobalamin (TC) deficiency (OMIM #275350) is caused by mutations in the *TCN2* gene[19]. Patients have elevations of both MMA and homocysteine, with normal serum Cbl levels due to circulating Cbl being predominantly bound to haptocorrin and not TC. Clinically, TC deficiency manifests in infancy with failure to thrive, defective humoral and cellular immunity, pancytopenia and megaloblastic anemia[20]. The hematological findings may result in misdiagnosing these patients with leukemia[21] or Cbl/folate deficiency[22]. The diagnosis can be confirmed by measuring the synthesis of TC in the patient's cultured fibroblasts[23]. If treatment with weekly IM Cbl injections is initiated early enough, the patient's symptoms

are reversible. However, to avoid the symptoms reappearing, lifelong compliance is required[20].

1.2.2.4 Haptocorrin Deficiency

Haptocorrin deficiency is caused by mutations in the *TCN1* gene[24]. Mutations in *TCN1* cause low serum Cbl levels, but the cellular metabolism is not compromised since Cbl must be bound to TC to be endocytosed by the cells. The function of haptocorrin is still unclear[25]. Haptocorrin deficiency is often asymptomatic and as such, is probably under-reported. However there are several cases of adults with mutations in haptocorrin that present with low serum Cbl, but no manifestations of deficiency including having reference levels of methylmalonic aciduria (MMAuria)[26-28]. This disorder, its effects and pathogenicity are still unclear[25].

1.2.2.5 Transcobalamin Receptor Deficiency

The Transcobalamin Receptor, encoded by the *CD320* gene[29] allows TC-bound cobalamin to enter the cell by receptor-mediated endocytosis. To date, only two mutations have been reported causing TCblR deficiency (OMIM#613646); c.262_264delGAG (E88del)[30] and c.297delA[31]. All patients have been picked up on newborn screen and are asymptomatic. The only detectable phenotype is their biochemical findings of isolated MMAuria or combined MMAuria and homocystinuria. All patients with TCblR deficiency have low cellular uptake of Cbl, measured by the uptake of [^{57}Co]CNCbl. However the activity of MCM ranges from

low to normal. There has been one patient reported with clinical findings in addition to the biochemical phenotype. He presented with blindness and central retinal artery occlusion, though whether this is an effect of the TCblR deficiency is unclear[32]. A study of an Irish cohort found the p.E88del homozygous in one control and heterozygous in 35 out of a total of 951 individuals giving a minor allele frequency of 0.02[33]. It is still unclear whether or not TCblR deficiency is a disease.

1.3 Intracellular Cobalamin Metabolism

In the cell, Cbl must be converted into its two cofactors, AdoCbl and MeCbl. The cellular metabolism of Cbl is summarized in Figure 3. The TC-Cbl complex is endocytosed, it enters the lysosomes where Cbl is released. Cbl then travels across the lysosomal membrane into the cytoplasm. This process is not yet well understood, but it is known to require two lysosomal membrane proteins: LMBD1 and ABCD4[34,35]. In the cytoplasm, Cbl binds to MMACHC and is then dealkylated or decyanated[36-38]. This process causes the reduction of cob(III)alamin to cob(II)alamin. Next, MMADHC is thought to function as a branchpoint[39], sending the Cbl to the mitochondria via a putative mitochondrial transporter or keeping it in the cytoplasm. In the cytoplasm, the cob(II)alamin is then methylated with a methyl group from MeTHF forming MeCbl by MS and methionine synthase reductase (MSR). MS then transfers this methyl group to homocysteine, forming methionine. In the mitochondria, cobalamin adenosyltransferase (ATR) along with an

unknown cobalamin reductase converts cob(II)alamin to cob(I)alamin. An adenosyl group from adenosine triphosphate (ATP) is added and AdoCbl is formed. ATR interacts with MCM, transferring AdoCbl to MCM which catalyzes the conversion of methylmalonylCoA into succinylCoA for use in the Krebs cycle[40]. Another protein, MMAA, is necessary for the association of AdoCbl with MCM and to protect the MCM-bound AdoCbl[1].

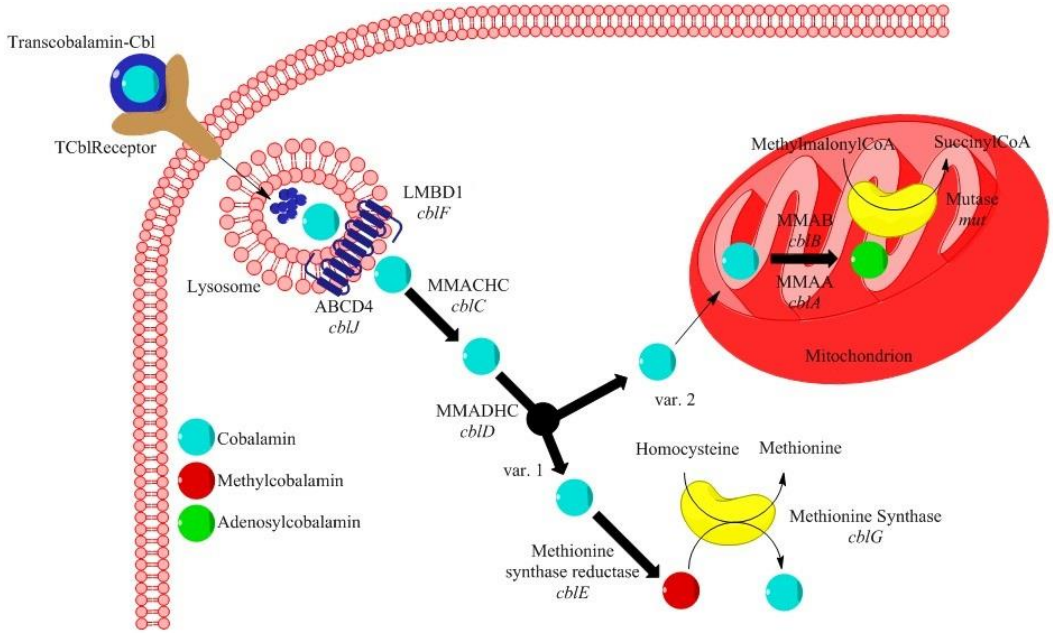

Figure 3: Intracellular Cobalamin Metabolism

Cobalamin is endocytosed into the lysosome, and then exported into the cytoplasm where it is converted into methylcobalamin and used for methionine synthase, or transported into the mitochondria where it is converted into adenosylcobalamin and used for methylmalonylCoA mutase. The proteins are labelled with their corresponding complementation groups underneath.

1.3.1 Inborn Errors of Cellular Cobalamin Metabolism

To date, nine complementation groups have been described in Cbl metabolism, along with their causal genes[41]. The resulting disorders are extremely rare and exhibit an autosomal recessive mode of inheritance[3].

These disorders typically present in infancy. They are often ascertained through newborn screening, as they result in the presence of measurable metabolites in the blood and urine. This most commonly occurs with a detection of elevated C3 acylcarnitine by tandem mass spectrometry[42]. However, some centers, such as the Quebec newborn screening program directly measures MMA in urine samples using thin layer chromatography[43]. Newborn screening programs allow the detection of asymptomatic patients, resulting in earlier treatment and prevention of the onset of any clinical symptoms in some patients[44].

1.3.1.1 Cellular Diagnosis

Somatic cell studies are used to evaluate enzyme function of MCM and MS, determining if one or both of these enzymes are affected by the defect. The function of these enzymes can be indirectly measured using radioactive precursor molecules and measuring the level of incorporation of the label into cellular macromolecules. [^{14}C]propionate is used to assay MCM[45] and [^{14}C]MeTHF to assay MS[46]. These incorporations are also measured in the presence of added OHCbl which helps narrow down which gene is defective[46]. The distribution of Cbl can also be examined by incubating cells in TC-bound [^{57}Co]-CNCbl. Using high performance liquid chromatography (HPLC) the synthesis of AdoCbl, MeCbl and OHCbl can be measured[46]. Low levels of AdoCbl or MeCbl point to a defect in the production of the cofactor. If the level of Cbl taken up by the cell is low this could mean a defect in uptake or retaining the Cbl in the patient's cells[41].

The expected results of these tests will be briefly discussed for each complementation group. If either propionate or MeTHF incorporation is lower than the reference range, somatic cell complementation is performed to determine which complementation group the patient belongs to. This assay will be fully explained below.

1.3.1.2 Somatic Cell Complementation Analysis

At the McGill vitamin B_{12} laboratory, we have a panel of cells belonging to all nine known complementation groups. To be able to perform the somatic cell complementation analysis, the [^{14}C]propionate or [^{14}C]MeTHF incorporations in the patient's fibroblasts must be low. These cell lines can then be fused with the unknown line with polyethylene glycol (PEG) forming multinucleated heterokaryon cells. As a control, the two cell lines are mixed together in the absence of PEG to give a baseline value[45]. If after fusion there is an increase of propionate incorporation, complementation is said to have occurred. This means that the two fused cell lines have a different defective gene in the pathway since in combination, a functional copy of both genes is produced. A depiction of this assay can be seen in Figure 4.

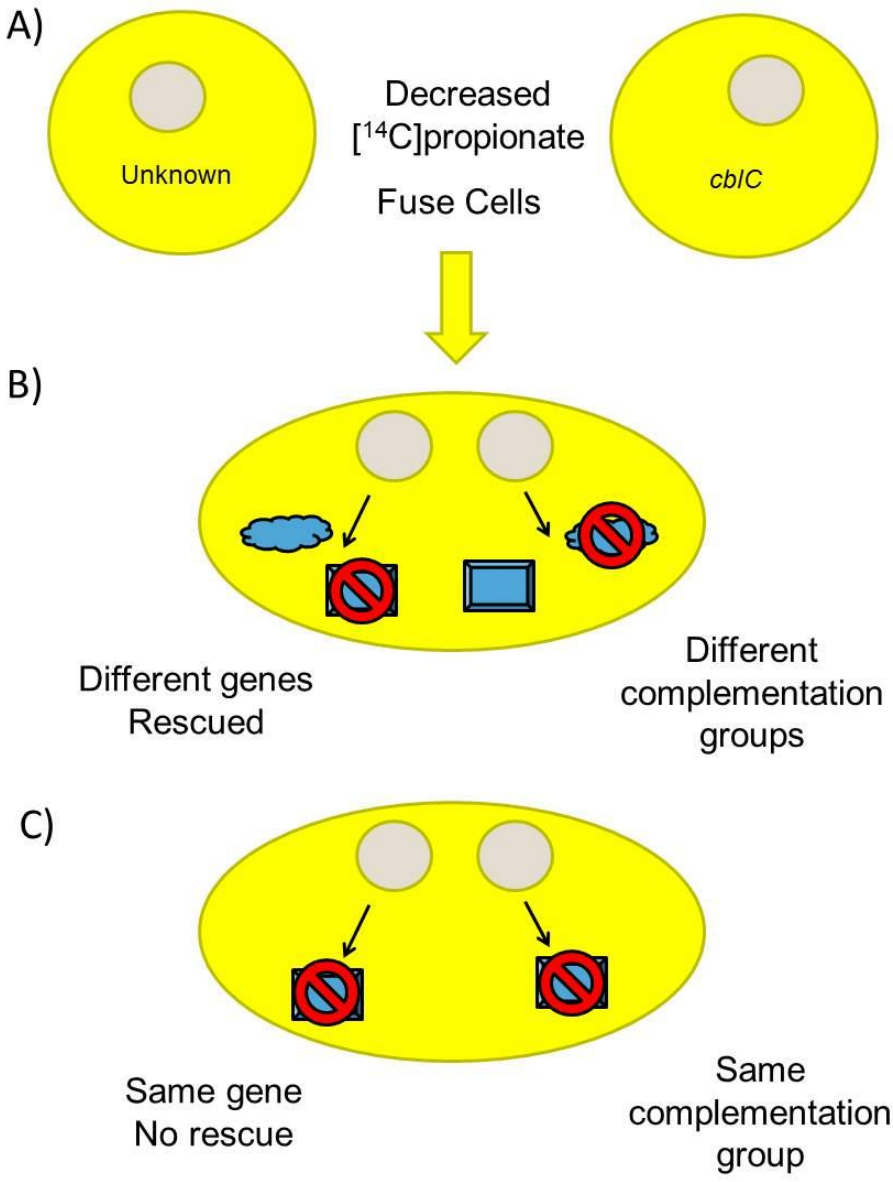

Figure 4: Somatic Cell Complementation Analysis

A) The unknown cell line with the biochemical phenotype is fused with a cell line of a known complementation group. There are two possible outcomes, depicted in B) and C). In B) there is a rescue in phenotype as the defect in both fused cell lines is in different genes. As such, both functional protein products are produced and the cell lines belong to different complementation groups. In C) there is no rescue in phenotype as the defect is in the same gene. No functional protein product is produced and the cell lines belong to the same complementation group.

The complementation test is not perfect. It can fail due to intragenic complementation, when two mutants with mutations in the same gene complement one another. This can happen if the mutations are in two different functional domains that act as separate functional elements and can rescue one another. For example, if two mutations are in different domains of a homomultimeric protein and can assemble to form an enzyme with partially corrected activity[47].

Second-site non-complementation occurs when mutations at two loci can interact and produce a mutant phenotype in double heterozygotes, even though they have mutations in separate genes. This can occur if the mutant proteins interact to produce non-functional complex that interferes with the wild-type form, if the mutant form of one protein sequesters the wild type form of the other in question resulting in a depletion, or if the simultaneous reduction of both protein product levels results in an insufficiency leading to a phenotype[47].

There have been previous examples from our laboratory of the complementation test giving misleading results. Initially, variant 2 of the *cblD* complementation group was thought to be a distinct subtype, *cblH*, as it did not complement with *cblA*, *cblB* or *mut* fibroblast lines and did not complement a panel of 28 *cblA* lines[45,48]. However, had the cells been fused with classic *cblD* lines, a lack of complementation would have been observed. This was a problem with the experimental interpretation rather than the complementation assay itself. There is also interallelic

complementation in *cblA* and *mut* patients, as not all patients within these classes are capable of complementing one another[45,49]. If a patient's fibroblast line's [^{14}C]propionate incorporation or [^{14}C]MeTHF incorporations are too elevated to perform complementation analysis then this can also give misleading results. We recently described a patient who could not be diagnosed by this method; however gene sequencing revealed two mutations in the *MMAB* gene, meaning this patient did in fact belong to the *cblB* complementation group. This patient likely has a mild biochemical phenotype and thus cannot be diagnosed by this method[50].

The complementation test is useful but has its drawbacks. Transformation rescues are considered to be the gold standard for assigning mutants or patients into groups[47]. They consist of inserting the wild type copy of the gene of interest into the mutants and looking for a rescue in phenotype. If the inserted gene and the mutated gene have very similar functions, and can replace one another, this insertion may correct the phenotype even though they are different genes. Also, if the gene product is very dependent on dosage, insertion of an endogenous product can have negative effects in its own. It is important to define the genetic mutations responsible in conjunction with the complementation test to avoid being misled.

1.3.1.3 Methylmalonic Aciduria

Decreased [^{14}C]propionate incorporation can be due to the absence of AdoCbl, or a block in MCM itself[1]. There are three complementation

groups in this category: *cblA* (OMIM #251100), *cblB* (OMIM #251110) and *mut* (OMIM #251000). They are caused by mutations in the *MMAA* (MMAA protein)[51], *MMAB* (ATR)[52] and *MUT* (MCM)[53] genes respectively. Patients from these complementation groups are characterized by MMAuria, and low levels of AdoCbl, though certain *mut* patients can have normal AdoCbl.

cblA patients often present in infancy with symptoms such as lethargy, failure to thrive, respiratory distress and vomiting[54]. They tend to have Cbl-responsive MMAuria, resulting in higher survival rates and better clinical outcomes than the other MMA complementation groups[55]. Regardless of response to therapy, late onset renal and neurological symptoms have been reported[56]. *cblB* patients show similar presentation as *cblA*, however only about half of patients are responsive to Cbl therapy which results in more severe clinical outcomes and decreased survival[54].

Patients with defects in the Cbl-dependent enzyme MCM can be classified into two subtypes; patients with residual enzyme function are *mut* $^{-}$ and those with no enzyme function are *mut*^{0} [57]. Phenotypic variability is high, though *mut*^{0} tend to be much more severely affected and present within the first week of life[54]. Patients may present with a severe phenotype of acidosis, multiorgan failure and death[58], with milder episodes of acidosis[59] or no clinical symptoms[60]. Chronic kidney disease is often seen[61] along with neurocognitive impairment which tends to be more severe in earlier onset patients[62]. *Mut* patients do not respond to therapy[54]

due to the defect being in the enzyme itself and not in the synthesis of the AdoCbl cofactor.

There are other causes of MMAuria outside of the metabolism of Cbl. There have been patients with mutations in *MCEE*, the gene encoding the enzyme methylmalonylCoA epimerase, which converts D-methylmalonic acid into the L- isoform useable by MCM (OMIM #251120)[63,64]. Patients exhibit mild MMAuria and a mildly reduced [^{14}C]propionate incorporation, though the clinical phenotype is variable. Some patients are asymptomatic while others have presented with hypotonia, ataxia, metabolic acidosis and/or seizures[65,66]. This variability in phenotype is even seen across patients with the same *MCEE* genotype[65]. The lack of a uniform clinical phenotype is consistent with previous observations that methylmalonylCoA spontaneously racemizes without the enzyme epimerase[67].

MethylmalonylCoA is converted into succinylCoA, and is subsequently used by the Krebs cycle enzyme succinate-CoA ligase. Thus, a defect in succinate-CoA ligase can cause mild elevations in MMA. Two genetic disorders have been identified affecting proteins that make up subunits of this enzyme. Mutations in *SUCLA2* (OMIM #612073), one of the two β subunits, primarily result in a mitochondrial DNA depletion syndrome[68]. Patients are typically healthy at birth but present at age 3-6 months with hypotonia and muscle weakness[69]. They have variable levels of lactic acidosis, encephalomyopathy, dystonia and deafness[70]. Mutations

in the α subunit *SUCLG1* (OMIM #245400) can have a much more severe phenotype with fatal infantile lactic acidosis and mitochondrial DNA depletion with patients passing away within the first week of life[71]. However later onset patients have been described presenting at 3 months[72] and 5 months[73]. Patients with different phenotypes have been reported such as those with liver impairment and no mitochondrial DNA depletion[73].

There have also been patients described with combined malonic aciduria and MMAuria (CMAMMA), with MMA excretion being higher than malonic acid excretion. This disorder has been shown to be caused by mutations in the methylmalonyl-CoA and malonyl-CoA synthetase gene *ACSF3* (OMIM #614265)[74,75]. Patients have variable clinical presentation. Some have adult-onset neurological manifestations such as seizures and cognitive decline, while others have a childhood presentation with ketoacidosis, failure to thrive and microcephaly[74]. Asymptomatic patients detected by newborn screening have also been described[75].

1.3.1.4 Homocystinuria

Low [^{14}C]MeTHF incorporation is caused by an absence of MeCbl or impaired function of the MS enzyme[1]. There are two complementation groups that result in this phenotype: *cblE* (OMIM #236270), caused by mutations in the *MTRR* gene (MSR protein)[76] and *cblG* (OMIM #250940) caused by mutations in *MTR* gene (MS protein)[77]. Biochemically, these defects result in homocystinuria and low MeCbl levels.

cblE and *cblG* patients ultimately have deficient function of the MS enzyme. Clinically, they are very similar with findings such as homocystinuria, megaloblastic anemia and some form of developmental retardation[78]. They tend to respond well to Cbl therapy and have correction of the biochemical and some of the clinical symptoms such as megaloblastic anemia and improvements of the neurological findings[78]. However, there has been at least one severe case reported where a patient did not respond to Cbl treatment and ultimately died from of haemolytic uremic syndrome[79].

Genetic defects have been reported in the folate pathway enzyme MTHFR, which reduces 5,10-MeTHF into 5-MeTHF, the methyl donor for MS. The decreased availability of 5-MeTHF results in a decreased function of MS, and homocystinuria (OMIM #236250). Mild mutations in this enzyme, including the one caused by the polymorphism c.677C>T[80], result in slightly elevated levels of homocystinuria in folate-deficient individuals. The minor allele has been associated with increased frequency of many conditions including coronary artery disease and neural tube defects[81,82]. Severe mutations in *MTHFR* cause a more serious disease with a variable clinical phenotype. Patients may present in infancy with encephalopathy, seizures and hypotonia[83]. Those with higher enzyme activity tend to have later onset symptoms such as neurodegenerative disorders and ataxia[84].

Another genetic disorder causing classic homocystinuria is a defect in the enzyme cystathionine beta-synthase (CBS) (OMIM #236200)[85]. It should be noted that this is not an inborn error in Cbl metabolism, but rather part of the transulfuration pathway involving the interconversion of cysteine and homocysteine. CBS catalyzes the conversion of homocysteine to cystathionine. Patients are diagnosed with thromboembolic events, spinal osteoporosis, seizures, mental retardation or optic lens dislocation. Some infants are diagnosed by newborn screening if elevated methionine is detected[86,87]. A subset of patients are responsive to vitamin B_6 and are less clinically affected than non-responsive individuals, particularly with regards to their mental capabilities[86].

There has been one patient reported to date with mutations in the tri-functional folate pathway enzyme MTHFD1[88]. This patient had combined homocystinuria and MMAuria, though the reason for the MMAuria is still unclear, as the folate pathway is not known to overlap with the mitochondrial AdoCbl pathway. The patient only had decreased MeCbl and normal AdoCbl synthesis. She also presented with severe combined immunodeficiency, atypical haemolytic uremic syndrome and megaloblastic anemia[88].

1.3.1.5 Combined Methylmalonic Aciduria and Homocystinuria

The complementation groups resulting in low [^{14}C]propionate and [^{14}C]MeTHF incorporation along with low AdoCbl and MeCbl are *cblF*

(OMIM #277380), *cblJ* (OMIM #614857), *cblC* (OMIM #277400), and *cblD* (OMIM #277410)[41]. *cblF*, *cblJ*, *cblC* and *cblD* are caused by mutations in the following genes, with protein names given in parentheses: *LMBRD1* (LMBD1)[34], *ABCD4* (ABCD4)[35], *MMACHC* (MMACHC)[89] and *MMADHC* (MMADHC)[48] respectively. The *cblC* complementation group and function of the MMACHC protein will be discussed in detail in section 1.4.

MMADHC is predicted to function at the branch-point of the Cbl pathway, so interestingly, specific mutations in *cblD* patients can result in low [^{14}C]propionate incorporation, or [^{14}C]MeTHF incorporation, or both[90]. It can therefore result in MMAuria, or homocystinuria, or both, as well as low AdoCbl, or low MeCbl or both. Patients can have developmental delay, megaloblastic anemia, seizures and respiratory distress[48,90]. The age of diagnosis is variable with some patients presenting in infancy, while others present as teenagers[48].

cblF and *cblJ* are very similar as both result in a defect in Cbl being transported out of the lysosome. These cell lines have a high proportion of CNCbl, representing unmetabolized Cbl trapped in the lysosomes[34,35]. *cblF* patients can clinically present with failure to thrive, developmental delay, minor facial abnormalities such as pegged teeth, congenital heart defects and megaloblastic anemia[34]. Patients often have low levels of Cbl in the serum, thought to be due to LMBD1's role in the intestinal uptake of Cbl[91]. There are only three *cblJ* patients described to date[35,92]. Two were early onset, presenting in the neonatal period with megaloblastic anemia,

hypotonia and heart defects[36]. One was later onset at the age of eight years whose symptoms included hyperpigmentation, dizziness, headaches and a transient ischemic attack[35,92].

1.4 The *cblC* Inborn Error of Cobalamin Metabolism

The *cblC* complementation group (OMIM #277400) is the most common inborn error of cobalamin metabolism with over 550 cases known worldwide[1]. The patients have impaired synthesis of both MeCbl and AdoCbl resulting in decreased function of both MCM and MS and combined MMAuria and homocystinuria[89].

cblC patients are often diagnosed by newborn screen with an elevated propionylcarnitine (C3) on tandem mass spectrometry, which suggests elevations of either propionylCoA or methylmalonylCoA[93]. Some systems also screen directly for MMA using thin-layer chromatography[94]. However, not every patient is picked up by newborn screen. If a patient's clinical picture is suggestive of *cblC*, a metabolic analysis measuring serum Cbl can be performed to rule out a nutritional deficiency. Levels of MMA, homocysteine and methionine along with an acylcarnitine profile are also necessary for a diagnosis. Once the biochemical phenotype is shown to be consistent with *cblC*, molecular diagnosis by mutation analysis of the *MMACHC* gene or somatic cell complementation analysis can be performed.

1.4.2 Clinical Symptoms

There is a great deal of clinical heterogeneity in the *cblC* disease with regards to severity and the systems affected. *cblC* has two phenotype classes: early onset and late onset[95].

Early onset patients present within the first year of life with clinical symptoms. In general, they are severely affected and have pathogenesis of multiple systems[89]. Patients can also present prenatally with intrauterine growth retardation, microcephaly and congenital heart disease[96]. The infantile presentation is most common, where patients often initially have feeding difficulties, failure to thrive, hypotonia and seizures[96]. Patients may later develop metabolic acidosis and neurological phenotypes such as developmental delay, seizures and hydrocephaly. Multiple hematological abnormalities are seen, particularly megaloblastic anemia but also thrombocytopenia, neutropenia, leukopenia or pancytopenia[95]. Renal failure is sometimes seen and is sometimes also accompanied by an atypical hemolytic uremic syndrome[97]. Some patients have an ocular phenotype which can include nystagmus, ophthalmological deterioration and pigmentary retinopathy which is unique to *cblC*[89].

Early onset patients have a much poorer clinical prognosis than later onset patients. Many die early on due to failure of one or more systems, typically between one and five months. If they survive into childhood they will likely suffer from neurological impairment[89]. Rosenblatt

et al. found only one child out of 44 early onset *cblC* patients was neurologically intact later on in childhood[95].

Late onset patients present after one year. The onset of symptoms is typically later in childhood, however the disorder can present in the teenage years or even in adulthood. These patients typically have milder clinical features than early onset patients. The oldest known *cblC* patient presented at 57 years (unpublished data). They are previously healthy but then often present with a neuropsychiatric phenotype such as confusion, dementia, psychosis[98], a decline in cognitive function sometimes with extrapyramidal signs[95]. Less commonly, thrombosis has been seen in some patients[98]. Spinal cord degeneration results in many symptoms such as numbness of extremities, gait disorders, lower extremity weakness and incontinence[96]. Since this form of *cblC* is not well known, particularly in adult medicine, diagnosis can be delayed[98,99]. There have been recently described cases of very atypical presentation of *cblC.* One described five children presenting with pulmonary arterial hypertension and renal thrombotic microangiopathy with no traditional signs of *cblC*. However, all had undiagnosed *cblC* confirmed by metabolic analysis and sequencing of *MMACHC*[100].

1.4.3 Outcome and Treatment

Early treatment of *cblC* patients is essential to avoid multi-systemic disease progression. However, diagnosis is challenging since the age of presentation and clinical phenotypes are quite variable[96]. *cblC* patients

can be treated with IM injections of OHCbl. Injections should be given daily to avoid reappearance of symptoms or worsening of the patient's condition[101-103]. Additional treatments include betaine, which is a methyl donor that can remethylate homocysteine to methionine independent of MS, along with folinic acid and a low protein diet. The use of the low protein diet is controversial[96]. CNCbl injections are not effective for *cblC* patients and its use rather than treating with OHCbl injections is associated with poor patient outcome[104]. *In vitro* studies show that the inability of cells to use CNCbl but to respond to OHCbl seems to be due to an inability of certain mutant forms of MMACHC to process the cyano group, whereas binding to OHCbl is unaffected[105]. OHCbl is likely useful as a therapeutic agent as it has been shown in *cblC* cell lines to reduce the number of reactive oxygen species[106]. It has also been shown to stabilize wild type MMACHC protein and to a lesser extent a late onset mutant MMACHC[107].

Compiled data on patient outcomes seem to show that early detection and treatment result in decreased mortality and improved weight gain in newborns[104]. Late onset patients that have a daily OHCbl injection, along with betaine supplementation seem to have the best clinical outcome, with improvement of the neurological and haematological findings. However, there have not been any clinical trials for any therapy for *cblC*[104].

1.4.4 *MMACHC* Gene

The *MMACHC* gene is located on chromosome 1p34.1[89]. It has five exons, but only the first four are coding. There is a small 5' untranslated region (UTR) and a large 3'UTR. The promoter of the gene, defined by acetylated lysine 27 on H3 histones[108], spans 1kb upstream and 1.3kb downstream from the ATG translation start site.

1.4.4.1 *MMACHC* Genotypes in *cblC* Patients

There are currently 81 known mutations that cause *cblC* disease. The most common, accounting for around 40% of alleles, is the c.271dupA that results in a frameshift and premature stop codon[109]. There appears to be a genotype phenotype correlation, with some mutations in homozygosity causing late onset disease: c.394C>T (p.R132X) and c.482G>A (p.R161Q). Other severe mutations in homozygosity are known to cause early onset symptoms: c.271dupA (frameshift) and c.331C>T (p.R111X)[110]. Allelic expression may partially explain this difference in phenotype. The severe c.271dupA and c.331C>T mutations are significantly underexpressed compared to wild type alleles, whereas c.394C>T and c.482G>A are overexpressed compared to the c.271dupA alleles. This also suggests that the latter two transcripts escape nonsense-mediated decay, resulting in higher levels of *MMACHC* transcript, and likely more MMACHC protein with residual function and a milder phenotype[110].

1.4.4.2 MMACHC Function

The crystal structure of the human MMACHC protein has been solved[111]. It was shown to have a large cavity for binding Cbl in the base-off conformation and an N-terminal flavodoxin nitroreductase domain. This suggests that it requires flavin in order to process Cbl. MMACHC catalyzes the glutathione-dependent dealkylation of AdoCbl and MeCbl and flavin-dependent decyanation of CNCbl[36-38]. Immunofluorescence and subcellular fractionation has shown that MMACHC is localized to the cytoplasm[112]. MMACHC has been shown to interact with the MMADHC protein by phage display and surface plasmon resonance experiments[39,113,114] and to interact with the cytoplasmic enzyme MS by co-immunoprecipitation and proximity ligation assay[115].

Residues towards the C-terminus of MMACHC showed similarity to the gram-negative bacterial protein TonB[89]. TonB is involved in the transport of iron and Cbl across the outer bacterial membrane[116] using a proton motive force[117] and an interaction with outer membrane transport proteins[118].

Expression studies done in day 11.5 mouse embryos showed that RNA expression of *Mmachc* is tissue specific, whereas expression of *Mmadhc* was ubiquitous. *Mmachc* was found to be highly expressed in the developing heart, lung, and cardiovascular and nervous systems[119].

1.5 HCFC1 and THAP11: Transcriptional regulators

Part of this project examined patients with a *cblC* phenocopy caused by transcriptional dysregulation of *MMACHC*. Most of these

patients had mutations in the X-linked transcriptional coregulator, *HCFC1* and one had mutations in its binding partner, the transcription factor *THAP11* which are necessary for transcriptional activation of *MMACHC*.

1.5.1 Eukaryotic Transcriptional Regulation

Transcription is regulated in eukaryotes by a variety of proteins including transcription factors, chromatin remodellers and kinases[120]. Protein coding genes are transcribed by RNA polymerase II (RNA pol II) in the nucleus. Transcription can be regulated during transcription, capping of the nascent mRNA and re-initiation of transcription[121]. Developmental processes and maintenance of homeostasis rely upon precise patterns of gene expression that are tightly regulated at multiple levels. For transcription to proceed, the DNA must be in an open, euchromatic conformation which can be achieved through histone modifications such as acetylation and methylation of lysine 4. RNA polymerase must be positioned at the core-promoter along with its general transcription factors.

Regulation of transcription is the most important factor in coordinating gene expression. Regulation occurs at multiple levels, one being the cis-regulatory elements. These are sequences of DNA on the same chromosome as the gene of interest. These elements include the core promoter where RNA pol II binds, promoter-proximal elements which are directly upstream the promoter and elements such as enhancers, silencers and insulators which can be several kilobases (kb) from the gene. Enhancers recruit transcription factors to their short and specific

DNA sequences[122] or chromatin modifying factors, such as the SWI/SNF complex. SWI/SNF functions by dissociating DNA from histones and promoting the binding of necessary transcription factors and recruiting other epigenetic modifying factors. Transcription factors bind short and specific DNA sequences of 6-12 base pairs (bp) and can have activating or repressive functions. If the summation of conditions is favourable for transcription, RNA pol II and the other general transcription factors bind the TATA box forming the pre-initiation complex. The DNA is unwound and elongation begins.

1.5.2 Transcriptional Regulators and Mendelian Disease

Mutations in transcription factors have been known to cause Mendelian disease. A well-known example is Rett Syndrome, a severe neurological disorder caused by dominant, loss of function mutations in the X-linked gene *MECP2*[123]. This gene encodes methyl CpG-binding protein 2. It binds to methylated DNA and induces transcriptional repression. It has also been implicated in autism and other neurodevelopmental disorders[124]. The *Mecp2*-null mouse model has shown that this factor is important in neuronal maturation and post-mitotic maintenance of neurons[125]. The exact mechanism by which mutations in *MECP2* cause disease is still unclear, though it seems to be due to incomplete repression of its targets in the absence of functional MECP2[126]. Gene expression studies have failed to identify many common dysregulated genes between patients[126].

Another example of transcriptional regulation involved in Mendelian disease is maturity onset diabetes of the young (MODY). MODY is a monogenic, autosomal dominant form of diabetes. It is characterized by genetic and clinical heterogeneity, with clinical symptoms manifesting in the first 25 years of life with defects in insulin secretion and pancreatic β-cell function[127]. Certain MODY subtypes are caused by mutations in genes encoding transcriptional regulators: MODY1[128], MODY3[129], MODY4[130], MODY5[131], MODY6[132], MODY7[133] and MODY9[134]. Transcriptional regulation is essential for normal pancreatic development, β-cell maintenance and glucose homeostasis[135]. These transcriptional regulators are thought to work in a complex regulatory network with multiple levels. Some transcription factors, such as *TCF1* (*HNF1A*), which causes MODY3, show evidence in mouse models of regulating *HNF4A, IPF1, NEUROD1 AND SHP1* in β-cells[136]. Others have a more simplistic model, such as NEUROD1, which causes the rare type MODY6. NEUROD1 regulates the insulin gene (*INS*) by binding directly to the promoter to regulate its expression[137]. The mouse model supports this hypothesis, with findings of diabetes, inadequate insulin expression[138] and defects in differentiation of pancreatic endocrine cells[139].

Another thanatos-associated protein (THAP) transcription factor, *THAP1* causes a dominant form of dystonia, *DYT6*, with 60% penetrance[140]. Patients have early onset of symptoms, which include dystonia of the craniocervical muscles and speech difficulties[141]. The

exact mechanism of the transcriptional dysregulation is unknown. Mutations seen in patients were engineered in a cell line. The mutant THAP1 showed decreased DNA binding, supporting that it is transcriptional dysregulation of THAP1 targets causing the clinical phenotype in patients[140].

Mutations in the transcriptional regulator *ATRX* can cause disorders such as Carpenter-Waziri syndrome[142], Juberg-Marsidi syndrome[143], Holmes-Gang syndrome[144], Smith-Fineman-Myers syndrome[145], Chudley-Lowry syndrome[146], X-linked mental retardation with spastic paraplegia[147] or X linked α-thalassemia with mental retardation (ATR-X syndrome)[148]. ATR-X syndrome is of particular relevance to this project as it is transcriptional dysregulation that causes a well-characterized monogenic disorder, α-thalassemia. However, on top of the hematological findings, it has additional clinical abnormalities such as developmental delay, severe mental retardation, no speech, characteristic facial abnormalities, genital abnormalities and mild skeletal deformities particularly affecting the fingers and toes[149]. Interestingly, 10% of patients have no α-thalassemia or any other hematological abnormality[150]. The ATRX protein is a helicase/ATPase that has the ability to remodel chromatin. Patients exhibit reduced levels of α-globin mRNA, therefore ATRX is presumably necessary for activation of its expression[151]. This is the only gene known to date that is dysregulated and leads to a clinical phenotype in ATR-X patients[152]. The additional clinical features remain unexplained. The

current hypothesis for the additional clinical findings is that ATRX is regulating specific target genes rather than functioning as a global transcriptional regulator[153]. In patients, reduced methylation is observed at repetitive sequences such as rRNA clusters, heterochromatic and subtelomeric repeats[154] and GC-rich variable number tandem repeats (VNTRs). The α-thalassemia locus has an upstream VNTR and expression of this gene is it is associated with the VNTR length. This explains hematologic phenotypic variability between patients, as patients with larger VNTRs more downregulation of the α-thalassemia locus[152,155].

1.5.3 HCFC1

HCFC1 (host cell factor 1, HCF-1) is a transcriptional co-regulator and not a transcription factor, as it lacks DNA-binding activity. It was first identified in herpes simplex virus transcription as being part of the VP16-induced regulatory complex, which acts as a switch promoting gene expression resulting in lytic infection of the virus[156]. It is an unusual protein, as it is synthesized as a 2000 amino acid polypeptide[157], but then migrates to the nucleus where it is cleaved repeatedly near its middle, resulting in a mixture of amino and carboxy terminal subunits that remain noncovalently associated[158]. Its role in regulating transcription is still very poorly understood[159]. It is known to be located at 5400 active CpG islands in HeLa cells and has three different sequence motifs, each associated with a different transcription factor that HCFC1 binds: ZNF143/THAP11, GABP and YY1[159]. Though it does not bind DNA directly, its Kelch repeat

domain allows it to associate with chromatin[160]. HCFC1 is a transcriptional intermediary that coordinates DNA-binding and chromatin-modifying factors, which ultimately results in the regulation of transcription. HCFC1 has been shown to recruit a variety of transcriptional activators or repressors, depending on the DNA sequence motifs of the gene that it is regulating. These regulators include histone methyltransferase SETD1A[161], histone demethylase PHF8[162] and the MOF histone acetyltransferase complex[163]. In coordination with these various factors, HCFC1 regulates a diverse set of processes such as spindle formation, DNA replication and cell division in HeLA cells[159] as well as protein synthesis and cellular metabolism in mouse embryonic stem cells[164]. When HeLa cells were treated with siRNA for *HCFC1,* the levels of 19% of all cellular mRNA transcripts changed compared to control cells, showing HCFC1 has a major regulatory effect in this cell type[159].

Dysregulation of *HCFC1* has been previously implicated in human disease. A promoter mutation disrupting a YY1 transcription factor binding site in the *HCFC1* sequence resulted in the over expression of *HCFC1*. This was found to be the cause of non-syndromic X-linked mental retardation, MRX3, in multiple members of an affected family[165]. The range of intellectual impairment ranged from mild to moderate and some males had mild facial abnormalities. However no consistent clinical features were seen throughout the family. Aggression, autistic behaviours and psychosis were also noted in some patients. Female carriers were

normal and showed high skewing of X-inactivation (>96%) favouring the wild type allele in lymphocytes[165].

In a cancer study, patients whose tumours over-expressed *HCFC1* had a poorer clinical outcome than those with normal expression, suggesting its role as a possible tumour driver[166]. This is not surprising based on its inferred role promoting the cell cycle[161,167,168].

1.5.4 THAP11

THAP11 (Ronin) is a small, one exon transcription factor that binds DNA in a site-specific fashion[169]. It is one of twelve members of the THAP family of transcription factors that all have a zinc finger-containing THAP domain. There is evidence for the THAP domain arising from the domestication of a DNA transposon[170]. THAP11 functions with HCFC1 as co-regulators of transcription. Together they have been shown, by chromatin immunoprecipitation sequencing (ChIP seq) using an antibody for THAP11, to interact in activating transcription of genes at a highly conserved enhancer element in mouse embryonic stem cells, particularly of genes involved in metabolism, mRNA splicing and transcription initiation[164]. THAP11 is also essential for pluripotency and embryogenesis in mouse embryonic stem cells where it establishes global transcriptional repression, ensuring pluripotency is maintained[171]. Knocking out *Thap11* results in embryonic stem (ES) cells that do not grow or divide, and overexpressing it results in ES cells that divide without differentiating into different cell types. There is also evidence for THAP11 being a tumour

driver as its expression was found to be up-regulated in primary colon cancer tumours and metastases[169]. Conflicting results were obtained in a study where the authors found that THAP11 was downregulated in many tumours including those of liver, vulva, thyroid, skin, pancreas and kidney, however it was upregulated in some lung and ovarian tumours they examined. They also noted that overexpressing *THAP11* in HeLa cells, fibroblasts and a breast cancer line resulted in decreased colony formation, reduced DNA synthesis and decreased expression of cell proliferation factor and known oncogene c-Myc[172]. This led the authors to hypothesize that THAP11 functions as a growth suppressor, conflicting with Dejosez[171]. The mechanisms of regulating stem cell and somatic cell proliferation are quite different, which could explain this discrepancy. Thap11 is only expressed in ovary and testes in adult mice and seems to be mostly important in early development, which could also explain the difference[173]. *Thap11* knock-out mice show a pre-implantation lethality and defects in their inner cell mass[171].

THAP11 is a member of a family of proteins that contains a highly conserved THAP domain at their N-terminus. This THAP domain contains a C2CH zinc finger motif. Humans have 12 THAP proteins (THAP0-THAP11). Eight out of twelve have an HCFC1 binding motif next to a coiled-coil domain. At their C terminus, there is a nuclear localization signal and a coiled-coil domain that allows for dimerization. The structure of THAP11's THAP domain has recently been solved by nuclear magnetic

resonance (NMR)[174] (Protein Data Bank ID: 2LAU). It displays the shared feature of the THAP domain, four invariant residues that coordinate with the zinc ion and a βαβ conserved fold around the zinc-coordination site.

1.6 Rationale and Objectives of Study

By 2006, our clinical laboratory had accumulated cultured fibroblasts from 339 *cblC* patients that had been diagnosed based on somatic cell complementation. DNA from 245 of these patients was used to show that mutations in the *MMACHC* gene caused *cblC* and all were subjected to mutation analysis by Sanger sequencing. At least one causal mutation in *MMACHC* was detected in all except 12 of the patients. Exome sequencing of one of these 12 patients by our collaborators at the University of Colorado, led by Dr. Tamim Shaikh, revealed a missense mutation, c.344C>T (p.Ala115Val) in the *HCFC1*[175] gene on the X chromosome. Sanger sequencing of *HCFC1* in the remaining 11 *cblC* samples found mutations in nine, including four additional missense mutations in exons 2 and 3.

Two DNA samples that were sent to Colorado still had no mutations after sequencing *MMACHC* and *HCFC1*. Upon re-examination of the biochemical studies, one patient had a high level of intracellular cobalamin that remained as unmetabolized cyanocobalamin. This phenotype is typically seen in the *cblF* and *cblJ* inborn errors. Sequencing of the *LMBRD1* gene showed homozygosity for the common c.1056delG mutation, meaning the patient actually belongs to the *cblF*

complementation group. In the remaining patient, genes that were known to interact with HCFC1 during transcriptional regulation were sequenced. The patient ended up having a homozygous c.240C>G (p.F80L) missense mutation in *THAP11*.

The first objective of this thesis was to sequence *MMACHC* in all remaining *cblC* patients that have one or no mutations in *MMACHC*. The second objective consisted of characterizing the *cblX* inborn error through analysis of clinical and metabolic data. The third aim was to characterize the patient with a homozygous mutation in *THAP11*. Finally, since HCFC1 and THAP11 are predicted to be global transcriptional regulators, all patients with mutations in *HCFC1* or *THAP11* were subjected to RNA sequencing for whole transcriptome profiling.

Mutations in transcriptional regulators have been known to cause Mendelian disorders for many years. However, until the present study there had been no known mutations described resulting in a metabolic disease caused by transcriptional dysregulation.

CHAPTER 2
Materials and Methods

2.1 Patient cells used in this study

All patient used in this study were diagnosed as *cblC* by somatic cell complementation analysis. By 2006, our clinical laboratory had accumulated cultured fibroblasts from 245 *cblC* samples that had been diagnosed based on somatic cell complementation. These cell lines all had decreased [^{14}C]propionate and [^{14}C]MeTHF incorporations, as well as low AdoCbl and MeCbl synthesis. DNA from all these patients was analyzed by Sanger sequencing of the *MMACHC* gene as part of the gene identification study[89]. Causal mutations in *MMACHC* were detected in all except 12 of the patients, which were used in this study.

Presently, there are a total of 339 *cblC* patients diagnosed by our clinical laboratory using somatic cell complementation analysis. Seventy-four of these patients had not been subjected to Sanger sequencing. Our clinical laboratory performs restriction enzyme digestion in order to identify common mutations in patients. This technique identified both causal mutations in 38 patients, leaving 36 patients with one or no causal mutations in *MMACHC* identified. These 36 patients were subjected to *MMACHC* sequencing in this study.

2.2 Cell Culture

Patient fibroblast cell lines from skin tissue biopsy were obtained from the Repository for Human Mutant Cells at the Montreal Children's Hospital. Cells were grown in minimum essential medium plus non-

essential amino acids (Gibco) with additional 5% fetal bovine serum (Intergen) and 5% iron enriched calf serum (Intergen). Cells transfected with pBABE and pLXSH vectors were maintained in medium with 1µg/mL puromycin (InvivoGen) and 100µg/mL hygromycin (Calbiochem), respectively.

2.3 DNA Extraction, Primer Design, PCR and Sequencing

DNA was extracted from fresh or frozen fibroblasts using the Puregene kit according to the manufacturer's instructions (Qiagen). DNA quality was assessed on a Nanodrop spectrophotometer (Thermo Fisher Scientific).

Primers were designed using the primer design software PrimerBlast. These primer pairs' nearest neighbour melting temperature had to be within 3°C of each other and not form any hairpin structures[176]. The specificity of these primers was checked using nBLAST[177] and the binding site was checked for the presence of single nucleotide polymorphisms (SNPs) using UCSC Genome Browser[178]. A list of the primers for this project, along with their polymerase chain reaction (PCR) conditions can be found in Appendix A.

PCR was performed using Taq or HotStar Taq (Qiagen). Each experiment had a positive control consisting of a DNA of known quality and a negative control using no DNA. The PCR products were visualized on 1.5% agarose gels with ethidium bromide staining and run with the GelPilot 100 bp Plus Ladder (Qiagen).

PCR products were sequenced using Sanger Sequencing at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada).

2.4 RNA Sequencing

Cell lines for RNA sequencing studies were grown in T75 tissue culture flasks until confluent. To ensure uniform cell treatment in order to minimize colluding effects on gene expression, cells were plated at a density of 9 million cells per flask and allowed to grow for five days. Cells were visually inspected at this point, ensuring confluency, regular cell shape and no abnormalities. If any significant abnormalities were noted the cultures were replated. If the cultures looked well, they were harvested, pelleted and dissolved in 2mL of Tri-Reagent (Sigma) and immediately frozen at -80ºC. The suspensions were transported on dry ice to the McGill University Genome Quebec Innovation Center where the RNA was extracted sequencing libraries were prepared as described below.

Total RNA from the suspended cell pellets was extracted using the miRNAeasy Mini Kit (Qiagen). Quality was assessed using the Agilent 2100 Bioanalyzer and quantified on the Nanodrop (Thermo Scientific). The RNA seq library was prepared using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina RS-122-2301) including Ribo-Zero Gold depletion to remove ribosomal RNA. The library quality was tested using Agilent 2100 Bioanalyser on the DNA-1000 chip.

The samples were indexed and four samples per lane were run on the Illumina Hi-Seq2000 sequencer, using 100bp paired end reads. The bioinformatic analysis was performed by Dr. Veronique Adoue of the Pastinen laboratory. The raw sequencing reads were trimmed for quality with a phred quality score between 33≥30 and a length greater than 32bp. The Illumina adapters were removed using Trimmomatic v0.22[179]. The filtered reads were aligned to the reference hg19 human genome using the programs TopHat v1.4.1[180] and Bowtie v0.12.8[181]. The duplicate reads were removed using Picard's MarkDuplicates.jar v1.80 (http://picard.sourceforge.net). The Fragment per Kilobase of exon model per Million mapped fragments (FPKM) values for the UCSC and Ensembl genes and transcripts were obtained using Cufflinks v2.1.1[182]. FPKM normalizes the abundance of reads from a given gene for the transcript length and library size. The reads were counted using HTSeq-Count[183]. Next, DESeq[184] and edgeR[185] were used to determine differential gene expression between patients and controls. Differences in gene expression between patients and three age and sex-matched controls were identified in normalized RNA-seq using ANOVA.

2.5 Restriction Digest Analysis of Common c.344C>T Mutation

Genomic DNA was amplified by PCR for *HCFC1* exon 3. 15µL of PCR product was digested by 2µL of *BspMI* restriction endonuclease (New England Biolabs) in the manufacturer's buffer overnight at 37ºC, followed by a 20 minute incubation at 65ºC to denature any remaining

enzyme. 25µL of the digestion product were run on a 1.2% agarose gel at 80V until the desired separation was achieved. The c.344C>T mutation creates a cut site for the enzyme. Patients had two bands, one at 321bp and the other at 150bp. Wild type individuals had a single band corresponding to the uncut PCR product at 471bp. Mutation carriers had all three bands, though the shortest 150bp band was not always visible.

2.6 Statistical Analyses

The independent two-way Mann-Whitney test was used to determine if there are statistically significant differences in the cellular biochemical data between *cblX* patients and *cblC* patients. This non-parametric test was used as opposed to the parametric t-test because the distributions of the data sets used were not normal and therefore in violation of one of the assumptions of the t-test. Statistical significance was deemed to be $p \leq 0.05$. The Mann-Whitney test was performed using the software found at http://vassarstats.net/utest.html.

CHAPTER 3
Results

3.1 Sequencing of *MMACHC* in samples diagnosed as *cblC*

At the beginning of this project, there remained 48 patients diagnosed as *cblC* by somatic cell complementation that had either one or no identified mutations in *MMACHC*. Twelve of these patients had been sequenced as part of a previous study[89]. 37 *cblC* patients had not been sequenced but had been subjected to clinical restriction digest analysis for common mutations in *MMACHC*: c.271dupA, c.331C>T, c.294C>T, c.328_331delAACC, c.388_390delTAC, c.440G>A, c.481C>T, c.482G>A and c.615C>G. Complete gene sequencing of *MMACHC* was undertaken, with the aim of discovering novel mutations in *MMACHC*. If there were still no mutations found, this patient would be considered a candidate for *cblX*, *THAP11* or perhaps even another complementation group. Sequencing results of confirmed *cblC* patients can be seen below in Table 1. The additional thirteen patients that had no mutations in *MMACHC* will be discussed below in section 3.1.2. The mutation analysis of the novel mutations is presented in Table 2.

Table 1: Summary of results from Sanger sequencing of the *MMACHC* gene in *cblC* patients.

Patient	Mutation 1	Mutation 2
WG3300	c.394C>T	c.394C>T
WG3308	c.328_331delAACC	c.328_331delAACC
WG3387	c.271dupA	c.271dupA
WG3408	c.271dupA	c.271dupA
WG3423	c.271dupA	c.608G>A
WG3431	c.271dupA	c.271dupA
WG3441	c.328_331delAACC	c.566G>A
WG3568	c.578T>C	c.578T>C
WG3601	c.482G>A	?
WG3638	c.271dupA	c.158T>C
WG3697	c.271dupA	c.578T>C
WG3720	c.328_331delAACC	c.292C>T
WG3725	c.271dupA	?
WG3728	c.439G>A	c.457C>T
WG3737	c.578T>C	c.482G>A
WG3738	c.482G>A	?
WG3749	c.271dupA	c.600G>A
WG4314	c.551_554dupCTAC	c.551_554dupCTAC
WG4354	c.435_436delAT	c.435_436delAT

Patient	Mutation 1	Mutation 2
WG3756	c.471G>A	c.471G>C
WG3813	c.567_568insT	c.482G>A
WG3838	c.80G>A	?
WG3857	c.271dupA	c.271dupA
WG3879	c.331C>T	c.589A>G
WG3902	c.271dupA	c.616C>T
WG3952	c.1A>T	c.1A>T
WG3964	c.609G>A	c.658_660delAAG
WG4018	c.271dupA	c.440G>C
WG4019	c.271dupA	?
WG4036	c.271dupA	c.389A>G
WG4080	c.271dupA	c.389A>G
WG4097	c.271dupA	?
WG4124	c.609G>A	?
WG3433	c.394C>T	?
WG3451	c.365A>G	c.457C>T
WG4152	c.158T>C	presumed regulatory
WG2417	c.271dupA	exon 2 deletion

Novel, unpublished mutations are highlighted in green. A question mark (?) denotes that a second mutation was not found.

3.1.1 Mutation Analysis of Novel Mutations in *MMACHC*

Table 2: Novel mutations in *MMACHC*

Mutation	Amino Acid Change	Polyphen-2 Score	Mutation Taster Score	dbSNP MAF
c.566G>A	p.R189H	1.000	0.995 Damaging	Not seen
c.158T>C	p.L53P	1.000	0.999 Damaging	Not seen
c.292C>T	p.Q98X	NA	1.000 Damaging	Not seen
c.551_554dupCTAC	p.R185YfsX3	NA	1.000 Damaging	Not seen

For each mutation, the predicted amino acid effect is given along with Polyphen-2 and Mutation Taster scores (all were predicted to be disease causing), minor allele frequencies (MAF) from dbSNP and the 1000 Genomes.

Both c.566G>A and c.158T>C missense mutations were predicted to be probably damaging by Polyphen-2 and Mutation Taster. Both of these amino acid residues were conserved until zebrafish. The nonsense mutation c.292C>T and the duplication c.551_554dupCTAC were predicted damaging by mutation taster. None of these mutations had been seen by the 1000 Genomes project.

3.1.1.1 Patients' Phenotype

The clinical details of the patients with the four novel mutations and presumed deletion and regulatory mutation can be seen in Table 3. The c.566G>A mutation was seen as compound heterozygous with the previously described c.328_331delAACC mutation. The patient was asymptomatic and only came to medical attention as a result of a positive newborn screen.

Only one mutation was identified in multiple patients. The c.158T>C (p.L53P) mutation was seen in two patients in this study, and has been found in an asymptomatic brother and sister pair, in combination with the c.271dupA mutation. Patient WG3638 had this c.158T>C allele in combination with the severe c.271dupA allele. She presented with weight loss, nausea and vomiting while she was pregnant at age 24. Patient WG4152 was only brought to medical attention at age 57 for an unrelated metabolic workup, due to complications from his squamous cell carcinoma of the tonsil. This patient also has a presumed regulatory mutation. Upon amplification of the patient's cDNA, his c.158T>C mutation that was heterozygous at the genomic level became homozygous. This is due to apparent allele dropout, likely due to lack of expression of one allele.

The c.292C>T (p.Q98X) mutation was seen in heterozygosity with the c.328_331delAACC, in a severely affected infant who presented with failure to thrive, hypotonia and seizures.

Patient WG2040 had a suspected deletion of exon 2. Upon amplification of the cDNA, two products were visible. Gel extraction and Sanger sequencing showed the shorter transcript was lacking exon 2. The resulting protein product of this deletion would likely be non-functional. The patient's second allele had the severe c.271dupA mutation. The patient had an early onset of disease and was very severely affected, with death occurring at one month of age.

The novel homozygous deletion c.551_554dupCTAC was seen in an asymptomatic baby that only came to clinical attention due to a positive newborn screen for MMA.

Table 3. Patients with novel mutations in *MMACHC* – Age of Onset and Phenotype

Patient	*MMACHC* variant 1	*MMACHC* variant 2	Age of Diagnosis	Clinical Description
WG3441	c.328_331del AACC	c.566G>A	NBS	Asymptomatic
WG3638	c.271dupA	c.158T>C	24 years (pregnancy)	Weight loss, nausea, vomiting
WG3720	c.328_331del AACC	c.292C>T	Birth	Failure to thrive, hypotonia, seizures
WG4152	c.158T>C	Presumed regulatory	57 years	Male with squamous cell carcinoma
WG2417	c.271dupA	Exon 2 deletion	Birth	Deceased 1 month
WG4314	c.551_554dup CTAC	c.551_554dup CTAC	NBS	Asymptomatic

The age of onset represents the age that the clinical symptoms began in affected patients. NBS=newborn screen. The clinical symptoms are taken from the patient description given by the physician at the time of referral, and as such, are not necessarily up to date.

3.1.2 Patients with no mutations in *MMACHC*

After Sanger sequencing of the *MMACHC* gene, there remained 13 patients with no mutations. Eleven had mutations in *HCFC1*, discussed in the next section. One, WG2110 had a homozygous missense mutation in *HCFC1*'s binding partner *THAP11*. This patient will be discussed in section 4. The last patient, WG2976 had no mutations in *MMACHC*, *HCFC1* or *THAP11*. This patient presented at 5 years of age with mild

developmental delay, macrocytosis and MMAuria and homocysteinemia. His serum Cbl was low-normal at 215 pg/mL (reference 200-600). The patient had multiple tooth abnormalities, as did other members of his maternal grandmother's family. He had a history of unexplained fevers, recurrent illnesses and dermatologic abnormalities. During his delivery he had a large blood loss requiring a transfusion, and he had anemia for the following three months. For the first year of life he had poor feeding, hypotonia and was fed through a G-tube. Cellular studies on his fibroblasts showed decreased [^{14}C]propionate and [^{14}C]MeTHF incorporation. There was very elevated proportion of total Cbl as CNCbl with 78.5%, low AdoCbl and low MeCbl. Somatic cell complementation results were difficult to interpret; fusion with a *cblC* line had poor complementation, it did not complement with a *cblF* line but the increase was greater when compared to the *cblC*. The cellular studies from this patient supported the possibility that this line did not belong to the *cblC* group. Sequencing of the *LMBRD1* gene showed the patient was homozygous for the common mutation, c.1056delG, and therefore belongs to the *cblF* group.

3.2 Characterization of patients with the *cblX* inborn error

Twelve male patients that had no mutations in *MMACHC* after the 2006 Sanger sequencing study were sent to Dr. Shaikh's laboratory at the University of Colorado. This included the patient they had originally exome sequenced. After sequencing *HCFC1*, they found ten patients with

missense mutations. Dr. Venditti's group at the NIH reported a mutation in *HCFC1* in a patient, WG4173, that had been diagnosed as a *cblD* by our laboratory. These patients were described as belonging to the *cblX* group[175] and will be discussed in this section. Sanger sequencing of the *HCFC1* gene in patient WG3627 after mutation analysis described in section 3.1 found a novel mutation c.80G>A. This meant our laboratory had accumulated 12 *cblX* patients, 11 had been diagnosed as *cblC* and one as *cblD*. Since the disease is X-linked, all patients are male.

3.2.1 Clinical Features of *cblX* Patients

Patients showed common symptoms such as epilepsy, hypsarrythmia on EEG, choreoathetosis, severe developmental delay and mental retardation. Their clinical findings are summarized in Table 4. The age of onset for all patients was in infancy, and some had prenatal findings such as intrauterine growth retardation (IUGR). Normal homocysteine levels were seen in four out of ten patients whose doctors provided this information. Sequencing revealed that all mutations in *HCFC1* were missense and located at the 5' end of the gene in exons 1, 2 and 3.

Table 4. *cblX* clinical findings and mutations in *HCFC1*

Patient Number	Age of onset	Clinical Symptoms	*HCFC1* variant
WG3352 (Index case)	4 months	Infantile spasms with hypsarrythmia, gyral cortical malformation, severe developmental delay, microcephaly	c.344C>T p.Ala115Val
WG1664	4 months	Intractable epilepsy, severe developmental delay, microcephaly, FTT, choreoathetosis	c.344C>T p.Ala115Val
WG1760	prenatal	Congenital microcephaly, neonatal epilepsy, severe developmental delay, FTT, choreoathetosis	c.344C>T p.Ala115Val
WG1990	<3 months	Infantile spasms with hypsarrythmia, absent neurological development	c.344C>T p.Ala115Val
WG2040	2 months	Epilepsy, severe developmental delay, choreoathetosis, neutropenia	c.344C>T p.Ala115Val
WG2085	prenatal	IUGR, infantile spasms with hypsarrythmia, severe developmental delay, microcephaly	c.344C>T p.Ala115Val
WG3301	4 months	Epilepsy, hypotonia	c.344C>T p.Ala115Val
WG2831	2 weeks	Acute neurologic and metabolic decompensation, ketoacidosis with hyperammonemia, microcephaly, abnormal EEG, died in infancy	c.343G>A p.Ala115Thr
WG4173	5 months	Developmental delay, FTT, epilepsy, ballismus	c.218C>T p.Ala73Val
WG3399	5 weeks	Epilepsy, severe developmental delay, FTT	c.217G>A p.Ala73Thr
WG2049	prenatal	IUGR, congenital microcephaly, FTT, seizures, blindness, hearing impairment, slightly dysmorphic	c.202C>G p.Gln68Glu
WG3627	Infancy (<1 year)	Intractable seizures, infantile spasms, developmental regression, hypotonia, chorea	c.80G>A p.Gly25Asp

Mutations and Clinical Findings of all reported *cblX* patients from the McGill Referral Centre: “Age of onset” represents the estimated age when symptoms began based on the doctor’s reports. Clinical summaries are based on the doctor’s report when sending the patient cell lines and as such, are not up to date. FTT=failure to thrive

3.2.2 Biochemical Features of *cblX* Patients

Since it was noted that *cblX* patients have a milder biochemical phenotype with regards to excretion of MMA and homocysteine[175] it was of interest to compare their cellular biochemical phenotype. The cellular biochemical data for all *cblX* (n=12) and *cblC* (n=306) were compiled for a comparative analysis. The Mann-Whitney test was performed between the averages of the *cblX* and *cblC* patients. The [^{14}C]propionate incorporation, [^{14}C]MeTHF incorporation and percentage of Cbl as MeCbl were significantly different between *cblX* and *cblC* patients with p-values of 0.005, 0.01 and $6x10^{-4}$ respectively. The Mann-Whitney test statistic values for these were 2713, 2613 and 2874. There was no statistically significant difference between the two groups for percentage of Cbl as AdoCbl with a p-value of 0.63 and Mann-Whitney test statistic of 2141. A depiction of these results can be seen in Figure 5.

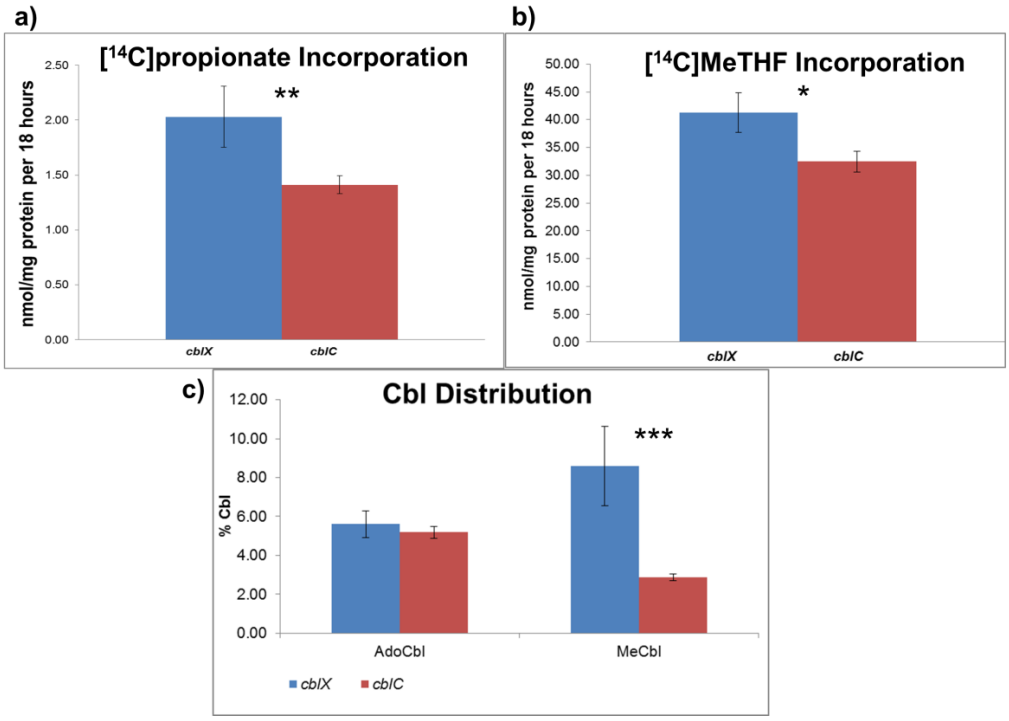

Figure 5: Biochemical Features of *cblX* versus *cblC* Patients

The average values of *cblX* (n=12) and *cblC* (n=306) for [^{14}C]propionate incorporation, [^{14}C]MeTHF incorporation and %AdoCbl and %MeCbl, compared using a Mann Whitney test. P-values were **0.005, *0.01, 0.30 and ***6x10-4 respectively. The bars represent the standard error. The reference ranges respectively are 10.8±3.7, 225±165, 15.29±4.2 and 57.95±6.7.

3.2.3 Haplotype Analysis of Patients with the Common c.344C>T Mutation in *HCFC1*

The c.344C>T mutation was seen in 7/12 samples from our laboratory and 9/15 *cblX* patients described to date. To determine if these patients share a common haplotype, the sequencing data from these patients was analyzed, evaluating common SNPs. The summary of this data is found in Table 5. Patients WG1990, WG2085 and WG3301 have informative variants that are different from the majority of samples.

Table 5. Haplotype Analysis of patients with the common c.344C>T mutation in *HCFC1*

Patient	Mutation	c.342+116	rs2071134 MAF 0.33 c.2283	rs730106 MAF 0.50 c.2841	rs3027875 MAF 0.29 c.5859
WG1760	c.344C>T	T	C	A	T
WG2040	c.344C>T	T	C	A	T
WG1990	c.344C>T	C	C	Failed	T
WG2085	c.344C>T	T	C	A	C
WG1664	c.344C>T	T	C	A	T
WG3352	c.344C>T	T	C	A	T
WG3301	c.344C>T	T	C	A	C

Three common SNPs in the *HCFC1* gene were chosen for analysis. One patient sample had an additional variant. The rsID for the SNPs, minor allele frequency (MAF) and position of the SNP is given in the above table. The sample marked as failed had a poor quality chromatogram for that exon and could not be read.

3.2.4 Carrier Testing of Mothers and Sisters of *cblX* Patients

As the disease is X-linked, this results in significant familial implications. Female relatives of the patient have two X chromosomes and are at risk of being carriers, with a 50% chance of passing on the X-linked condition to their sons and the mutant allele to their daughters. As such, family testing is extremely important. Maternal DNA and one sister's DNA was received for carrier testing from the referring physicians. The samples were received from WG3301's mother and maternal half-sister. We already had a stored fibroblast sample from WG1664's mother, with the ID number WG1722. All female relatives tested were carriers of the common c.344C>T mutation as determined by Sanger sequencing of exon 3 of the *HCFC1* gene.

3.2.5 Restriction Digest Analysis of the Common c.344C>T Mutation

We developed a restriction digest assay that can be performed in our molecular diagnostic lab to test for the mostly frequently identified *HCFC1* mutation, c.344C>T. This test allows clinical reports to be sent back to the referring physicians, and this information can be used in future prenatal diagnoses for female carriers, if they choose to take the test. This information can then also be shared with other female relatives such as aunts and cousins who can also choose to undergo carrier testing, as they are also at risk. The results of the restriction digest assay can be found in Figure 6.

The results of the restriction digest analysis correspond with those from the Sanger sequencing. Patients have a faint wild type band due to incomplete digestion. However, this does not affect the analysis as it is much fainter than the others and males only have one copy of the X-chromosome and can therefore not be misclassified as carriers.

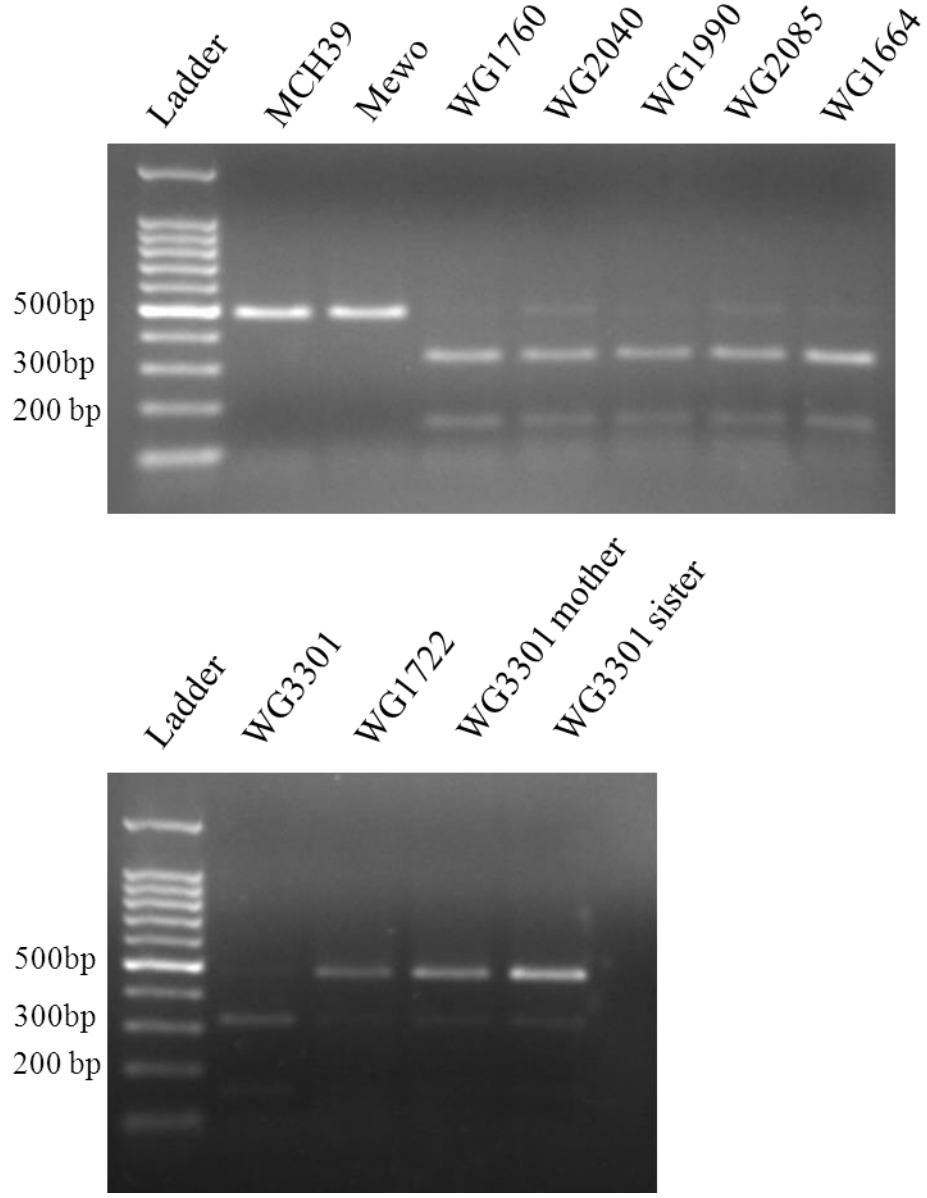

Figure 6: Restriction Digest Analysis of *HCFC1* c.344C>T

MCH39 and Mewo (+/+) only have the wild type band, WG1760 through WG3301 (-/-) are *cblX* patient samples have the mutant bands at 321bp and 150bp, though the wild type band is very faintly visible, WG1722 and WG3301 mother and sister (+/-) have all three bands, though the low 150bp is very faintly visible.

3.3 Characterization of a Patient with Mutations in *THAP11*

3.3.1 Clinical and Biochemical Features of Patient WG2110

One patient, WG2110, characterized as *cblC* by complementation analysis, had no mutations in either *MMACHC* or *HCFC1* by Sanger sequencing. Since his phenotype was more consistent with *cblX*, his sample was screened at the University of Colorado in Dr. Shaikh's laboratory for mutations in the gene encoding a zinc finger transcription factor that is well known to function with HCFC1 in transcriptional activation at the *MMACHC* locus: *THAP11*.

This child was born to unrelated Moroccan parents. At two months of age, the boy presented with seizures. At four months his overall development was delayed. At twelve months, a metabolic work-up found mild MMAuria. No homocysteine could be detected in his urine. Treatment was initiated and consisted of a protein restricted diet, IM hydroxocobalamin injections, pyridoxine, folic acid, and betaine. However, this did not prevent the progression of the child's clinical sequelae. He subsequently developed encephalopathy, tetraplegia, profound mental retardation, and seizures. Later, when the measurement of total plasma homocysteine became available, his levels were slightly elevated. The boy died at age ten and the parents did not consent to postmortem examination. The family declined subsequent prenatal diagnosis and has since had three normal children.

The patient's cellular studies showed decreased [^{14}C]propionate and [^{14}C]MeTHF incorporation. There was elevated CNCbl, low Cbl and

MeCbl though both were higher than normally seen in other inborn errors of Cbl metabolism. The results of the complementation analysis of the patient's cultured fibroblasts indicated a *cblC* diagnosis.

3.3.2 *THAP11* Mutation Analysis

Sequencing of *THAP11* identified a homozygous mutation, c.240C>G (p.F80L) located in the conserved zinc finger-containing THAP domain. The mutation is two residues away from P78, which is responsible for interacting with the zinc ion. F80L is predicted to be probably damaging by Polyphen-2 and the residue is conserved in 10/12 human THAP proteins (Figure 7) and is conserved throughout vertebrates down to the zebrafish orthologue (Figure 8).

THAP1	N	A	V	P	T	I	F	L
THAP2	D	A	V	P	T	I	F	D
THAP3	N	A	V	P	T	V	F	A
THAP4	T	A	V	P	S	I	F	H
THAP5	T	A	V	P	T	I	F	S
THAP6	G	V	I	P	S	I	F	D
THAP7	G	A	V	P	T	I	F	E
THAP8	D	A	V	P	S	I	F	S
THAP9	G	A	V	P	S	V	S	L
THAP10	G	A	V	P	T	L	H	R
THAP11	V	R	V	P	T	I	F	P
THAP0	N	A	I	P	T	I	F	D
Matches	0	0	10	11	8	8	9	0

Figure 7: Multi-sequence alignment of human THAP proteins

The aligned sequences of THAP0 through THAP11 were aligned using Consurf. The sequence of THAP11 is highlighted in yellow and the F80 residue is highlighted in red. The number of matches in other THAP proteins compared to THAP11 is listed below each amino acid (max=12/12). The F80 residue is conserved in 9/12 other THAP proteins with the differing S and H residues being polar while F is non-polar.

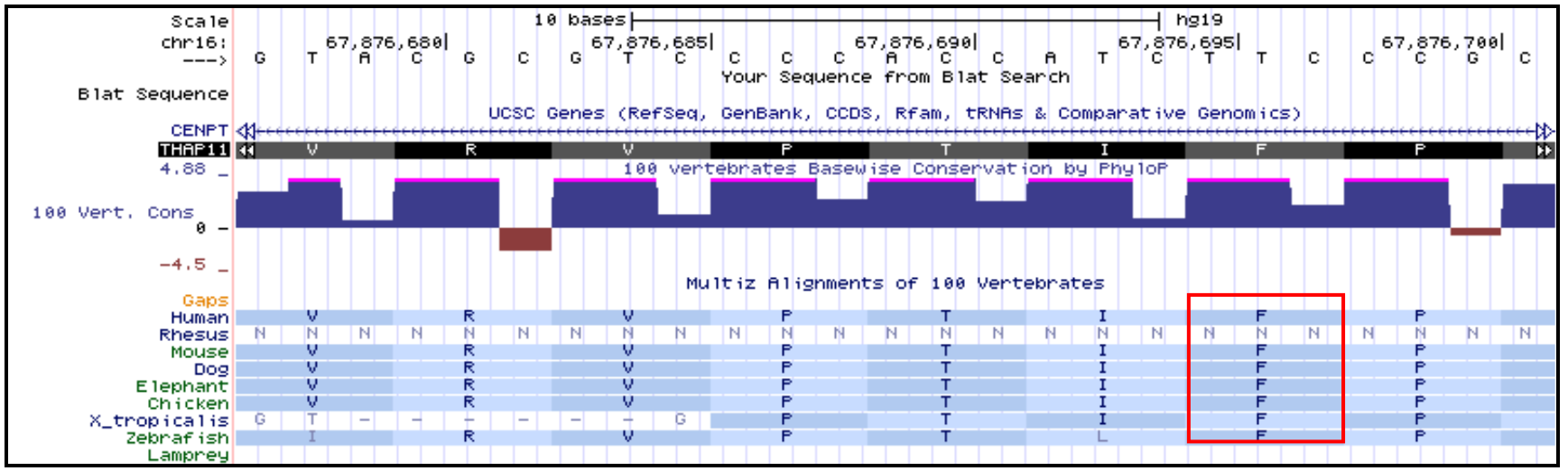

Figure 8: Multi-species alignment of THAP11

Starting at position V74, the multi-species alignment is shown generated in UCSC genome browser. The F80 residue is highlighted and conserved in all THAP11-containing species.

3.4 Transcriptome Analysis of *cblX* and *THAP11* Patient Fibroblasts

RNA was isolated from three control fibroblast lines, 11 *cblX* fibroblast lines with mutations in *HCFC1* and the patient fibroblast line with a homozygous mutation in *THAP11*. The control fibroblast lines were young males under the age of ten, whose cell lines were under 10 generations in culture. RNA sequencing (RNA Seq) was performed and the transcriptomes from each line were analyzed. Gene expression was measured for each sample using a unitless, normalized RNA-expression score to allow for comparison. Patients were compared to controls which resulted in statistically significant differential expression for five genes: *TMOD2*, *PIEZO2*, *ZNF883*, *NAT1*, *MMACHC*. Expression in all five genes was downregulated in the all of the patient samples. The expression of three genes *HCFC1*, *MMACHC* and *TMOD2* are shown below in Figure 9. The expression of *MMACHC* was undetectable in the following patient lines: WG2831, WG2049 and WG3352. Expression of both mutated genes, *HCFC1* and *THAP11* was not different between patients and controls.

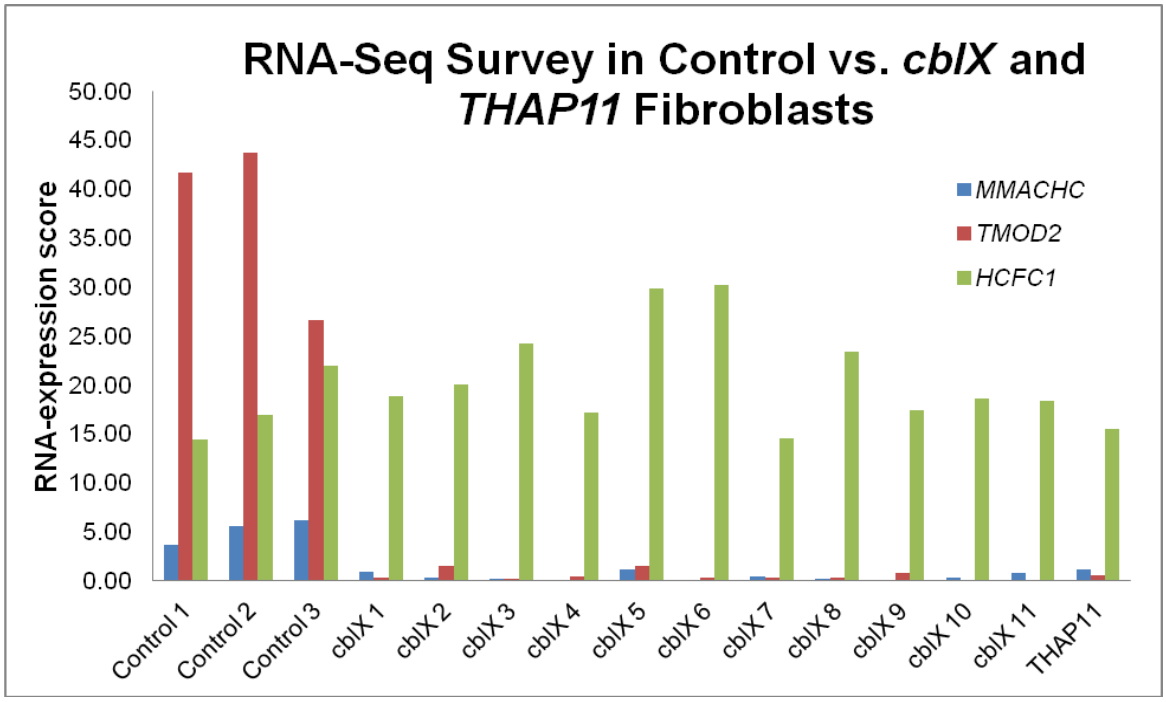

Figure 9: Results of RNA sequencing experiments in control and *cblX* fibroblasts

The above figure depicts expression for three genes (*MMACHC*, *TMOD2* and *HCFC1*) for each of the samples. *HCFC1* expression (green) was not different between patients and controls. Expression of *MMACHC* (blue) which causes the *cblC* inborn error and *TMOD2* (red) were much lower in the patients, with some lines having undetectable expression.

CHAPTER 4
Discussion

4.1 Novel Mutations in *MMACHC*

Of the 50 *cblC* patients diagnosed by somatic cell complementation that had one or no mutations in *MMACHC*, 37 were confirmed to have a homozygous or compound heterozygous mutation in *MMACHC*. Mutations in the transcriptional co-regulator *HCFC1* were identified in eleven of the 13 remaining patients and these patients were diagnosed as *cblX.* One patient had mutations in the transcription factor*THAP11* and one patient was homozygous the common *LMBRD1* mutation and was diagnosed as *cblF*.

Now all patients diagnosed as *cblC* in our laboratory have at least one pathogenic mutation in *MMACHC*. There are still eight *cblC* patients where one causal mutation has yet to be identified and two patients whose presumed mutations need to be characterized. WG4152 had a lack of one allele at the cDNA level as the c.158T>C mutation is heterozygous at the gDNA level and homozygous in the cDNA. No additional mutation has been found in the coding region, promoter or in the untranslated regions (see Alison Brebner Honour's Thesis, 2012). WG2417 had an observed absence of exon 2 at the cDNA level. These two samples, along with WG3838 have been sent to Baylor College of Medicine, where they have developed a gene panel, preferentially targeting all the known genes that cause inborn errors of Cbl metabolism or methylmalonic aciduria. This will check for missense mutations, however they also have the ability to

simultaneously search for large deletions and duplications that are missed by traditional Sanger sequencing. For those other remaining seven patients, perhaps they also have an additional mutation that cannot be picked up by Sanger sequencing the *MMACHC* gene. If the gene panel discovers novel mutations, we can then check for them in these patients as well. Another approach being developed is a long range PCR assay, which amplifies the entire *MMACHC* gene and allows the visualization of large deletions or duplications. RNA can also be extracted from these patients and converted to cDNA, allowing the detection of splice site mutations, or exon deletions such as in WG2417, as they cannot be picked up by Sanger sequencing.

Regarding the novel mutations in *MMACHC*, all are very likely to be causal and pathogenic. Both missense mutations, c.566G>A and c.158T>C have high Polyphen-2 and Mutation Taster scores. The affected residues are highly conserved, being identical in rhesus, mouse, dog, elephant, chicken, *Xenopus* and zebrafish. The c.292C>T and c.551_554dupCTAC are both predicted to result in a truncated protein product. The best characterized mutation is c.158T>C, which has been seen in four of our late onset patients and has since been published in a fifth patient[186]. Three were in combination with the severe c.271dupA allele in an asymptomatic sibling pair and a 24-year-old who presented during pregnancy. The fourth was in combination with the unknown mutation that causes an absence of one allele at the cDNA level in a 57

year old, who was picked up during a metabolic workup for complications during his squamous cell carcinoma. None have any reported neurological or ocular symptoms and are generally doing well. All other mutations were found in single patients. The c.566G>A was found in combination with c.328_331delAACC. The deletion appears to be a severe mutation[110,187], therefore it is likely that the c.566G>A is a mild mutation as the patient did not present with a severe phenotype that would be expected from having two very deleterious mutations. The nonsense mutation c.292C>T was also seen in combination with the c.328_331delAACC deletion. This appears to be a severe mutation as the infant presented with severe symptoms at birth. This is consistent with its effect on the protein as it results in a truncation at the 98th amino acid, quite early on in the protein likely resulting in very little residual function. Finally, the homozygous c.551_554dupCTAC appears to be a mild mutation as the infant was asymptomatic and only picked up due to newborn screening. However, the child is still young and perhaps the symptoms did not get a chance to develop due to the early diagnosis. The frameshift does occur quite late in the protein at the 185th amino acid. Perhaps with continual treatment the patient will remain healthy. For the three mutations each seen in only one patient, additional patients are required to confirm the mutations' effect on patient phenotype.

4.2 *cblX* and *THAP11* Patients

Compiling these patients allowed the observation that clinically, they are quite different from *cblC*. All *cblX* patients and the single *THAP11* patient have very severe neurological phenotypes. These are more severe than what is typically seen in *cblC* patients, though they share some common clinical features. *cblX* patients have presented with symptoms such as severe developmental delay, choreoathetosis and intractable seizures. Their neurological symptoms were often reported to not be reversed by OHCbl injections. However, biochemically they had a milder phenotype than most *cblC* patients. *cblX* patients had lower levels of metabolite excretion than *cblC* patients, with some having homocysteine levels within the reference range. The cellular studies also showed a milder phenotype. Both [^{14}C]propionate and [^{14}C]MeTHF incorporations were significantly higher in *cblX* patients versus *cblC* patients and there was a higher level of MeCbl that was very statistically significant with $p=6x10^{-4}$. This higher level of MeCbl is consistent with many *cblX* patients having milder homocysteine levels than *cblC* patients. This mild biochemical phenotype is likely due to residual MMACHC that is wild type versus *cblC* patients that have no wild type MMACHC. This also supports the patients' severe neurological phenotype being due to additional transcriptional effects on genes other than *MMACHC*, due to mutations in *HCFC1* or *THAP11*. Both these genes have been implicated in development and as important regulators of the cell cycle. Future experiments would characterize the effects of HCFC1 and THAP11 in

causing this severe neurological phenotype in patients. Animal models or inducing these mutations into human neural cell types and observing their ability to grow and differentiate would be interesting future directions.

Close parallels can be drawn with *HCFC1* and *THAP11* to *ATRX*, which causes ATR-X syndrome. This syndrome also results in a specific, previously characterized phenotype, α-thalassemia. Like *cblX* and *THAP11*, there is an additional severe neurological phenotype that includes craniofacial malformations, severe developmental delay, hypotonia and characteristic facial features[154]. ATR-X syndrome is due to a mutation in a transcription factor that regulates the α-globin locus, along with other genes presumably necessary in development. The perturbed regulation of these genes likely results in all the additional features seen in ATR-X patients. The same can be hypothesized for *cblX* and the *THAP11* patient. Dysregulation of *MMACHC* causes the biochemical phenotype and would likely contribute to the neurological findings. However, there are likely other genes important in neurological development that are dysregulated and result in the severe neurological findings in *cblX* patients.

Interestingly, there was a common mutation in the *HCFC1* gene seen in 9/15 *cblX* patients described to date. The haplotype analysis showed that the common c.344C>T mutation must have arisen multiple times. It is not likely that this is an ancient mutation that has remained on multiple haplotypes, as the disease would be more prevalent in the general

population, especially considering it is X-linked. The mutation arising sporadically is the likely scenario, since according to Hardy-Weinberg equilibrium, a third of mutant alleles are lost each generation when affected males do not reproduce. To keep the allele frequency constant one third of the alleles must arise through *de novo* mutations. The two informative variants, c.343+116 and rs3027875 show that the seven patients do not share the same haplotype and therefore the mutation has likely arisen independently multiple times. Three patients have received the mutation from a carrier mother and two additional patients have a reported maternal family history of infant deaths or developmental delay, though relatives have not been tested in these two families. The mutation arose sporadically in one of the subject's mother in Yu *et al.*177 as it was not seen in the maternal grandmother.

4.3 Lack of Complementation Between *cblC* and *cblX*

One major question that arises from this work is why fibroblasts from *cblC* and *cblX* patients do not complement one another. Fusion between the two cell lines should result in heterokaryons containing functional copies of both *MMACHC* and *HCFC1*. However, some mechanism is preventing the correction. This is consistent with preliminary results from retroviral transduction experiments, attempting to correct *cblX* and *THAP11* mutant fibroblasts by inserting wild type *HCFC1* and *THAP11* respectively. These experiments have so far failed to correct their respective mutant cell line's biochemical phenotypes. The *HCFC1*

transduction experiment was also performed in the laboratories of our collaborators, Charles Venditti (NIH) and Tamim Shaikh, using both the pcDNA mammalian expression vector and adenoviruses, both of which also failed to result in correction. These results are preliminary as the expression of the transduced gene has not yet been confirmed. However, transduction of wild type *MMACHC* was able to rescue both cell lines' [^{14}C]propionate incorporation.

Wysocka *et al.*[160] corrected an observable defect in *HCFC1* mutant cells using the same pBABE retroviral transduction system in HeLa cells. One possible explanation for the lack of complementation between *cblC* and *cblX* is that in this paper, correction did not occur until the cells divided. If the heterokaryons produced during the complementation analysis do not divide, or experience a growth disadvantage compared to unfused cells, this could explain the non-correction in this assay.

There are several possible explanations for these observations, all of which are presently theoretical and would need additional experimental support. Both HCFC1 and the THAP group of proteins dimerize before joining larger protein complexes[157,188]. It is possible that the mutant copies of HCFC1 and THAP11 are interfering with the ability of the wild type to correct the phenotype. Along similar lines, mutant HCFC1 and THAP11 could be occupying the *MMACHC* promoter and not allow the wild type proteins to bind.

Another possibility is that *HCFC1* and *THAP11* are translated near their “home” nucleus and imported immediately. This localization of nuclear RNAs has been seen before (Reviewed in Di Liegro *et al.*[189]). This would mean that wild type *HCFC1* would stay in the *cblC* nucleus and not be able to enter the *cblX* nucleus and result in the transcription of wild type *MMACHC*. Tagging the wild type *HCFC1* in the *cblC* cell line with green fluorescent protein (GFP) would allow us to use fluorescent microscopy to localize the tagged protein product at the cellular level. If it is not localized to both nuclei, then this is an obvious reason why the wild type HCFC1 would not correct the biochemical defect.

Ultimately, another possibility is that neither of these genes is causing the patients’ phenotype. However, the evidence for *HCFC1*, with 15 patients with damaging missense mutations in this gene, is quite strong. However, there is only one patient to date with mutations in *THAP11*. It is a very good candidate; biologically, it is a known to function with HCFC1. The mutation analysis is strong with a very damaging substitution at an extremely conserved residue among multiple THAP domain-containing proteins and across other species. The RNA seq results are also convincing, with very low expression of the same five downregulated genes as the *cblX* patients. Unfortunately, this patient passed away and his family is not interested in participating in research, therefore family material from the healthy unaffected siblings is not a possibility to rule out this candidate as being causal. ChIP sequencing

studies could help by giving further evidence, if THAP11 does not bind to the *MMACHC* promoter and other genes that are down-regulated in this patient, this would provide support that this patient is not transcribing *MMACHC* adequately due to a defect in *THAP11*.

4.4 RNA Sequencing Studies

The RNA sequencing studies also had surprising results. The effects of these mutations in *HCFC1* are very distinct as the specificity of the down-regulated genes was quite striking. There were not many genes dysregulated in patient samples. Only five genes in fibroblasts showed statistically significantly dysregulation between patients and controls. This is intriguing as HCFC1 is known to be a global transcriptional regulator with 19% of genes affected in HeLa cells when it is knocked down[159]. These transcriptome experiments need to be expanded to other cell types, particularly tissues affected such as neuronal cell types to see if these cells also have a specific dysregulation of a subset of genes.

The expression of *HCFC1* in *cblX* fibroblasts and *THAP11* in the *THAP11* patient was the same as control lines. This is expected as missense mutations are not expected to affect mRNA levels. Expression of most genes predicted to be regulated by *HCFC1* was not altered.

The most interesting gene knocked down in patients was *MMACHC,* which was much lower in patients and even absent in three *cblX* patients. This supports the results seen on qRT-PCR in the initial paper done on

two patients[175] and confirms why they cellularly mimic *cblC* patients, due to severely reduced, or even absence of MMACHC protein.

From this analysis, an interesting candidate that could be relevant to the patients' neurological phenotype was found. *TMOD2* was significantly down-regulated in patients compared to controls and was the top hit. TMOD2 is an actin-binding protein that is responsible for capping the pointed end[190]. It is highly expressed in the cerebellum, occipital lobe and slightly lower in the cortex and frontal lobe[191]. Its overexpression in rat PC12 cells, a model for neuronal differentiation, has been shown to inhibit neurite formation[192]. Decreased expression of *TMOD2* is of interest since it has been implicated as a regulator of neurite development and could be related to neurologic problems in *cblX* patients.

There were three other statistically significant downregulated candidates: *PIEZO2*, *ZNF883* and *NAT1.* PIEZO2 is a mechanosensitive ion channel. Gain of function mutations cause a subtype of Distal Arthrogryposis. Patients have limited eye movements and restricted lung disease[193]. ZNF883 is an uncharacterized zinc finger protein. NAT1 is a drug metabolizing enzyme of interest in pharmacogenetics and cancer research[194]. NAT1 acetylates its substrates, one of which includes the folate catabolite p-aminobenzoylglutamate[195]. Presently, none of these candidates have been deemed to be worth pursuing.

Currently, this RNA-seq data is being expanded by including other patients with different inborn errors of Cbl metabolism. This will allow

elucidation of whether or not transcriptional dysregulation is specific to *cblX* and the *THAP11* patient.

The *cblX* and *THAP11* patient represent novel genetic causes of inborn errors of Cbl metabolism. Their mode of causing disease through transcriptional dysregulation is novel for inborn errors of metabolism. Studying these patients can help answer many biological questions related to vitamin B_{12} metabolism and its transcriptional regulation, the biology of the transcription factors HCFC1 and THAP11, and their role in development and gene expression.

References

1. Watkins, D. & Rosenblatt, D. Inborn Errors of Cobalamin Absorption and Metabolism. *American Journal of Medical Genetics* **157**, 33-44 (2011).
2. Hodgkin, D.G. *et al.* The crystal structure of the hexacarboxylic acid derived from B_{12} and the molecular structure of the vitamin. *Nature* **176**, 325-8 (1955).
3. Froese, D.S. & Gravel, R.A. Genetic disorders of vitamin B metabolism: eight complementation groups--eight genes. *Expert Rev Mol Med* **12**, e37 (2010).
4. Gruber, K., Puffer, B. & Krautler, B. Vitamin B_{12}-derivatives-enzyme cofactors and ligands of proteins and nucleic acids. *Chemical Society Reviews* **40**, 4346-4363.
5. Yates, A.A., Schlicker, S.A. & Suitor, C.W. Dietary Reference Intakes: the new basis for recommendations for calcium and related nutrients, B vitamins, and choline. *Journal of the American Dietetic Association* **98**, 699-706 (1998).
6. Matthews, R.G. Cobalamin-dependent Methionine Synthase. in *Chemistry and Biochemistry of B_{12}* (ed. Banerjee, R.) (1999).
7. Stabler, S.P. & Allen, R.H. Vitamin B_{12} deficiency as a worldwide problem. *Annual review of nutrition* **24**, 299-326 (2004).
8. Herbert, V. & Zalusky, R. Interrelations of vitamin B_{12} and folic acid metabolism: folic acid clearance studies. *Journal of Clinical Investigation* **41**, 1263 (1962).
9. Stabler, S.P. Vitamin B_{12} deficiency. *The New England journal of medicine* **368**, 2041-2 (2013).
10. Kozyraki, R. & Cases, O. Vitamin B_{12} absorption: mammalian physiology and acquired and inherited disorders. *Biochimie* **95**, 1002-7 (2013).
11. Toh, B.H., van Driel, I.R. & Gleeson, P.A. Pernicious anemia. *The New England journal of medicine* **337**, 1441-8 (1997).
12. Andersen, C.B.F., Madsen, M., Storm, T., Moestrup, S.K. & Andersen, G.R. Structural basis for receptor recognition of vitamin-B_{12}-intrinsic factor complexes. *Nature* **464**, 445-448.
13. Majumder, S., Soriano, J., Louie Cruz, A. & Dasanu, C.A. Vitamin B_{12} deficiency in patients undergoing bariatric surgery: Preventive strategies and key recommendations. *Surgery for Obesity and Related Diseases* (2013).
14. Broch, H., Imerslund, O., Monn, E., Hovig, T. & Seip, M. Imerslund-Grasbeck anemia. A long-term follow-up study. *Acta paediatrica Scandinavica* **73**, 248-53 (1984).
15. Aminoff, M. *et al.* Mutations in *CUBN*, encoding the intrinsic factor-vitamin B_{12} receptor, cubilin, cause hereditary megaloblastic anaemia 1. *Nature genetics* **21**, 309-13 (1999).

16. Tanner, S.M. *et al.* Amnionless, essential for mouse gastrulation, is mutated in recessive hereditary megaloblastic anemia. *Nature genetics* **33**, 426-9 (2003).
17. Gordon, M.M. *et al.* A genetic polymorphism in the coding region of the gastric intrinsic factor gene (*GIF*) is associated with congenital intrinsic factor deficiency. *Human Mutation* **23**, 85-91 (2004).
18. McIntyre, O.R., Sullivan, L.W., Jeffries, G.H. & Silver, R.H. Pernicious Anemia in Childhood. *The New England journal of medicine* **272**, 981-6 (1965).
19. Li, N., Rosenblatt, D.S., Kamen, B.A., Seetharam, S. & Seetharam, B. Identification of two mutant alleles of transcobalamin II in an affected family. *Human Molecular Genetics* **3**, 1835-40 (1994).
20. Trakadis, Y.J. *et al.* Update on transcobalamin deficiency: clinical presentation, treatment and outcome. *Journal of inherited metabolic disease* (2013).
21. Schiff, M. *et al.* Should transcobalamin deficiency be treated aggressively? *Journal of inherited metabolic disease* **33**, 223-9 (2010).
22. Aitelli, C., Wasson, L. & Page, R. Pernicious anemia: presentations mimicking acute leukemia. *Southern medical journal* **97**, 295-7 (2004).
23. Rosenblatt, D.S., Hosack, A. & Matiaszuk, N. Expression of transcobalamin II by amniocytes. *Prenatal diagnosis* **7**, 35-9 (1987).
24. Lin, J.C., Borregaard, N., Liebman, H.A. & Carmel, R. Deficiency of the specific granule proteins, R-binder/transcobalamin I and lactoferrin, in plasma and saliva: a new disorder. *American Journal of Medical Genetics* **100**, 145-51 (2001).
25. Carmel, R., Parker, J. & Kelman, Z. Genomic mutations associated with mild and severe deficiencies of transcobalamin I (haptocorrin) that cause mildly and severely low serum cobalamin levels. *British journal of haematology* **147**, 386-91 (2009).
26. Carmel, R. & Herbert, V. Deficiency of vitamin B_{12}-binding alpha globulin in two brothers. *Blood* **33**, 1-12 (1969).
27. Carmel, R. A new case of deficiency of the R binder for cobalamin, with observations on minor cobalamin-binding proteins in serum and saliva. *Blood* **59**, 152-6 (1982).
28. Carmel, R. R-binder deficiency. A clinically benign cause of cobalamin pseudodeficiency. *JAMA : the journal of the American Medical Association* **250**, 1886-90 (1983).
29. Quadros, E.V., Nakayama, Y. & Sequeira, J.M. The protein and the gene encoding the receptor for the cellular uptake of transcobalamin-bound cobalamin. *Blood* **113**, 186-92 (2009).

30. Quadros, E.W. *et al.* Positive Newborn Screen for Methylmalonic Aciduria Identifies the First Mutation in *TCblR*/*CD320*, the Gene for Cellular Uptake of Transcobalamin-bound Vitamin B_{12}. *Human Mutation* **31**, 924-929 (2010).
31. Anastasio, N. *et al.* Mutations in *TCblR*, the gene for the transcobalamin receptor, result in decreased cellular uptake of vitamin B_{12} and methylmalonic aciduria. *Molecular genetics and metabolism* **98**, 122 (2009).
32. Karth, P., Singh, R., Kim, J. & Costakos, D. Bilateral central retinal artery occlusions in an infant with hyperhomocysteinemia. *Journal of AAPOS* **16**, 398-400 (2012).
33. Pangilinan, F. *et al.* Transcobalamin II receptor polymorphisms are associated with increased risk for neural tube defects. *Journal of medical genetics* **47**, 677-85 (2010).
34. Rutsch, F. *et al.* Identification of a putative lysosomal cobalamin exporter altered in the *cblF* defect of vitamin B_{12} metabolism. *Nature Genetics* **41**, 234-239 (2009).
35. Coelho, D. *et al.* Mutations in *ABCD4* cause a new inborn error of vitamin B_{12} metabolism. *Nature genetics* **44**, 1152-1155 (2012).
36. Kim, J., Hannibal, L., Gherasim, C., Jacobsen, D.W. & Banerjee, R. A Human Vitamin B_{12} Trafficking Protein Uses Glutathione Transferase Activity for Processing Alkylcobalamins. *Journal of Biological Chemistry* **284**, 33418-33424 (2009).
37. Kim, J., Gherasim, C. & Banerjee, R. Decyanation of vitamin B_{12} by a trafficking chaperone. *Proceedings of the National Academy of Sciences* **105**, 14551-14554 (2008).
38. Hannibal, L. *et al.* Processing of alkylcobalamins in mammalian cells: A role for the MMACHC (cblC) gene product. *Molecular Genetics and Metabolism* **97**, 260-266 (2009).
39. Gherasim, C., Hannibal, L., Rajagopalan, D., Jacobsen, D.W. & Banerjee, R. The C-terminal domain of CblD interacts with CblC and influences intracellular cobalamin partitioning. *Biochimie* **95**, 1023-32 (2013).
40. Banerjee, R., Gherasim, C. & Padovani, D. The tinker, tailor, soldier in intracellular B_{12} trafficking. *Current Opinion in Chemical Biology* **13**, 484-491 (2009).
41. Watkins, D. & Rosenblatt, D.S. Lessons in biology from patients with inborn errors of vitamin B_{12} metabolism. *Biochimie* **95**, 1019-22 (2013).
42. Sweetman, L. Newborn screening by tandem mass spectrometry (MS-MS). *Clinical chemistry* **42**, 345-6 (1996).
43. Auray-Blais, C., Cyr, D. & Drouin, R. Quebec neonatal mass urinary screening programme: from micromolecules to macromolecules. *Journal of inherited metabolic disease* **30**, 515-21 (2007).

44. Armour, C.M. *et al.* A patient with an inborn error of vitamin B_{12} metabolism (*cblF*) detected by newborn screening. *Pediatrics* **132**, e257-61 (2013).
45. Watkins, D., Matiaszuk, N. & Rosenblatt, D.S. Complementation studies in the *cblA* class of inborn error of cobalamin metabolism: evidence for interallelic complementation and for a new complementation class (*cblH*). *J Med Genet* **37**, 510-3 (2000).
46. Watkins, D. & Rosenblatt, D.S. Functional methionine synthase deficiency (*cblE* and *cblG*): Clinical and biochemical heterogeneity. *American Journal of Medical Genetics* **34**, 427-434 (1989).
47. Hawley, R.S. & Gilliland, W.D. Sometimes the result is not the answer: the truths and the lies that come from using the complementation test. *Genetics* **174**, 5-15 (2006).
48. Coelho, D. *et al.* Gene identification for the *cblD* defect of vitamin B_{12} metabolism. *The New England journal of medicine* **358**, 1454-64 (2008).
49. Raff, M.L., Crane, A.M., Jansen, R., Ledley, F.D. & Rosenblatt, D.S. Genetic characterization of a *MUT* locus mutation discriminating heterogeneity in mut^0 and mut^- methylmalonic aciduria by interallelic complementation. *The Journal of clinical investigation* **87**, 203-7 (1991).
50. Illson, M.L. *et al.* High resolution melting analysis of the *MMAB* gene in *cblB* patients and in those with undiagnosed methylmalonic aciduria. *Molecular genetics and metabolism* **110**, 86-9 (2013).
51. Dobson, C.M. *et al.* Identification of the gene responsible for the *cblA* complementation group of vitamin B_{12}-responsive methylmalonic acidemia based on analysis of prokaryotic gene arrangements. *Proc Natl Acad Sci U S A* **99**, 15554-9 (2002).
52. Dobson, C.M. *et al.* Identification of the gene responsible for the *cblB* complementation group of vitamin B_{12}-dependent methylmalonic aciduria. *Human Molecular Genetics* **11**, 3361-3369 (2002).
53. Ledley, F.D. *et al.* Mapping of human methylmalonyl CoA mutase (*MUT*) locus on chromosome 6. *Am J Hum Genet* **42**, 839-46 (1988).
54. Matsui, S.M., Mahoney, M.J. & Rosenberg, L.E. The natural history of the inherited methylmalonic acidemias. *New England Journal of Medicine* **308**, 857-861 (1983).
55. Hörster, F. *et al.* Long-term outcome in methylmalonic acidurias is influenced by the underlying defect (mut^0, mut^-, *cblA*, *cblB*). *Pediatric research* **62**, 225-230 (2007).

56. Watkins, D., Rosenblatt, D. & Fowler, B. Disorders of Cobalamin and Folate Transport and Metabolism. in *Inborn Metabolic Diseases* (eds. Saudubray, J.-M., Berghe, G. & Walter, J.) 385-402 (Springer Berlin Heidelberg, 2012).
57. Willard, H.F. & Rosenberg, L.E. Inherited deficiencies of human methylamlonyl CoA mutase activity: Reduced affinity of mutant apoenzyme for adenosylcobalamin. *Biochemical and Biophysical Research Communications* **78**, 927-934 (1977).
58. Stokke, O., Eldjarn, L., Norum, K., Steen-Johnsen, J. & Halvorsen, S. Methylmalonic Acidemia a new inborn error of metabolism which may cause fatal acidosis in the neonatal period. *Scandinavian Journal of Clinical & Laboratory Investigation* **20**, 313-328 (1967).
59. Shapira, S.K., Ledley, F.D., Rosenblatt, D.S. & Levy, H.L. Ketoacidotic crisis as a presentation of mild ("benign") methylmalonic acidemia. *The Journal of Pediatrics* **119**, 80-84 (1991).
60. Ledley, F.D., Levy, H.L., Shih, V.E., Benjamin, R. & Mahoney, M.J. Benign Methylmalonic Aciduria. *New England Journal of Medicine* **311**, 1015-1018 (1984).
61. Manoli, I. *et al.* Targeting proximal tubule mitochondrial dysfunction attenuates the renal disease of methylmalonic acidemia. *Proceedings of the National Academy of Sciences* **110**, 13552-13557 (2013).
62. O'Shea, C.J. *et al.* Neurocognitive Phenotype of Isolated Methylmalonic Acidemia. *Pediatrics* **129**, e1541-e1551 (2012).
63. Bobik, T.A. & Rasche, M.E. Identification of the Human Methylmalonyl-CoA Racemase Gene Based on the Analysis of Prokaryotic Gene Arrangements: Implications for Decoding the Human Genome. *Journal of Biological Chemistry* **276**, 37194-37198 (2001).
64. Dobson, C.M. *et al.* Homozygous nonsense mutation in the *MCEE* gene and siRNA suppression of methylmalonyl-CoA epimerase expression: a novel cause of mild methylmalonic aciduria. *Molecular genetics and metabolism* **88**, 327-33 (2006).
65. Gradinger, A.B. *et al.* Atypical methylmalonic aciduria: frequency of mutations in the methylmalonyl CoA epimerase gene (*MCEE*). *Human Mutation* **28**, 1045-1045 (2007).
66. Bikker, H. *et al.* A homozygous nonsense mutation in the methylmalonyl-CoA epimerase gene (*MCEE*) results in mild methylmalonic aciduria. *Human Mutation* **27**, 640-3 (2006).
67. Mazumder, R., Sasakawa, T., Kaziro, Y. & Ochoa, S. Metabolism of Propionic Acid in Animal Tissues IX. Methylmalonyl Coenzyme A Racemase. *Journal of Biological Chemistry* **237**, 3065-3068 (1962).

68. Elpeleg, O. *et al.* Deficiency of the ADP-forming succinyl-CoA synthase activity is associated with encephalomyopathy and mitochondrial DNA depletion. *The American Journal of Human Genetics* **76**, 1081-1086 (2005).
69. Ostergaard, E. *et al.* Mitochondrial encephalomyopathy with elevated methylmalonic acid is caused by *SUCLA2* mutations. *Brain* **130**, 853-861 (2007).
70. Carrozzo, R. *et al. SUCLA2* mutations are associated with mild methylmalonic aciduria, Leigh-like encephalomyopathy, dystonia and deafness. *Brain : a journal of neurology* **130**, 862-74 (2007).
71. Ostergaard, E. *et al.* Deficiency of the α subunit of succinate–coenzyme A ligase causes fatal infantile lactic acidosis with mitochondrial DNA depletion. *The American Journal of Human Genetics* **81**, 383-387 (2007).
72. Ostergaard, E. *et al.* A novel missense mutation in *SUCLG1* associated with mitochondrial DNA depletion, encephalomyopathic form, with methylmalonic aciduria. *European journal of pediatrics* **169**, 201-5 (2010).
73. Valayannopoulos, V. *et al.* New *SUCLG1* patients expanding the phenotypic spectrum of this rare cause of mild methylmalonic aciduria. *Mitochondrion* **10**, 335-41 (2010).
74. Sloan, J.L. *et al.* Exome sequencing identifies ACSF3 as a cause of combined malonic and methylmalonic aciduria. *Nature genetics* **43**, 883-6 (2011).
75. Alfares, A. *et al.* Combined malonic and methylmalonic aciduria: exome sequencing reveals mutations in the *ACSF3* gene in patients with a non-classic phenotype. *Journal of Medical Genetics* **48**, 602-605 (2011).
76. Leclerc, D. *et al.* Cloning and mapping of a cDNA for methionine synthase reductase, a flavoprotein defective in patients with homocystinuria. *Proc Natl Acad Sci U S A* **95**, 3059-64 (1998).
77. Li, Y.N. *et al.* Cloning, mapping and RNA analysis of the human methionine synthase gene. *Hum Mol Genet* **5**, 1851-8 (1996).
78. Watkin, D. & Rosenblatt, D.S. Functional methionine synthase deficiency (*cblE* and *cblG*): Clinical and biochemical heterogeneity. *American Journal of Medical Genetics* **34**, 427-434 (1989).
79. Palanca, D. *et al. cblE*-Type Homocystinuria Presenting with Features of Haemolytic-Uremic Syndrome in the Newborn Period. in *JIMD Reports - Case and Research Reports, 2012/5*, Vol. 8 (eds. Zschocke, J., Gibson, K.M., Brown, G., Morava, E. & Peters, V.) 57-62 (Springer Berlin Heidelberg, 2013).
80. Frosst, P. *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature genetics* **10**, 111-3 (1995).

81. Klerk, M. *et al.* MTHFR 677C>T polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA : the journal of the American Medical Association* **288**, 2023-31 (2002).
82. van der Put, N.M. *et al.* Mutated methylenetetrahydrofolate reductase as a risk factor for spina bifida. *Lancet* **346**, 1070-1 (1995).
83. Balasubramaniam, S., Salomons, G.S. & Blom, H.J. A case of severe methylenetetrahydrofolate reductase deficiency presenting as neonatal encephalopathy, seizures, microcephaly and central hypoventilation. *Journal of Pediatric Neurology* **11**, 135-140 (2013).
84. Goyette, P., Frosst, P., Rosenblatt, D.S. & Rozen, R. Seven novel mutations in the methylenetetrahydrofolate reductase gene and genotype/phenotype correlations in severe methylenetetrahydrofolate reductase deficiency. *American journal of human genetics* **56**, 1052-9 (1995).
85. Mudd, S.H., Finkelstein, J.D., Irreverre, F. & Laster, L. Homocystinuria: An Enzymatic Defect. *Science* **143**, 1443-5 (1964).
86. Mudd, S.H. *et al.* The natural history of homocystinuria due to cystathionine beta-synthase deficiency. *American journal of human genetics* **37**, 1-31 (1985).
87. Skovby, F., Gaustadnes, M. & Mudd, S.H. A revisit to the natural history of homocystinuria due to cystathionine beta-synthase deficiency. *Molecular genetics and metabolism* **99**, 1-3 (2010).
88. Watkins, D. *et al.* Novel inborn error of folate metabolism: identification by exome capture and sequencing of mutations in the *MTHFD1* gene in a single proband. *J Med Genet* **48**, 590-2 (2011).
89. Lerner-Ellis, J.P. *et al.* Identification of the gene responsible for methylmalonic aciduria and homocystinuria, *cblC* type. *Nature Genetics* **38**, 93-100 (2006).
90. Miousse, I.R. *et al.* Clinical and molecular heterogeneity in patients with the *cblD* inborn error of cobalamin metabolism. *The Journal of Pediatrics* **154**, 551-6 (2009).
91. Laframboise, R., Cooper, B.A. & Rosenblatt, D.S. Malabsorption of vitamin B_{12} from the intestine in a child with *cblF* disease: evidence for lysosomal-mediated absorption. *Blood* **80**, 291-2 (1992).
92. Kim, J.C. *et al.* Late onset of symptoms in an atypical patient with the *cblJ* inborn error of vitamin B_{12} metabolism: diagnosis and novel mutation revealed by exome sequencing. *Molecular genetics and metabolism* **107**, 664-8 (2012).

93. Chace, D.H., DiPerna, J.C., Kalas, T.A., Johnson, R.W. & Naylor, E.W. Rapid diagnosis of methylmalonic and propionic acidemias: quantitative tandem mass spectrometric analysis of propionylcarnitine in filter-paper blood specimens obtained from newborns. *Clinical chemistry* **47**, 2040-4 (2001).
94. Auray-Blais, C., Giguere, R., Paradis, D. & Lemieux, B. Rapid thin-layer chromatographic method for the detection of urinary methylmalonic acid. *Clinical biochemistry* **12**, 43-5 (1979).
95. Rosenblatt, D.S. *et al.* Clinical heterogeneity and prognosis in combined methylmalonic aciduria and homocystinuria (*cblC*). *Journal of Inherited Metabolic Disease* **20**, 528-538 (1997).
96. Carrillo-Carrasco, N., Chandler, R. & Venditti, C. Combined methylmalonic acidemia and homocystinuria, *cblC* type. I. Clinical presentations, diagnosis and management. *Journal of Inherited Metabolic Disease* **35**, 91-102 (2012).
97. Sharma, A.P., Greenberg, C.R., Prasad, A.N. & Prasad, C. Hemolytic uremic syndrome (HUS) secondary to cobalamin C (*cblC*) disorder. *Pediatric nephrology* **22**, 2097-103 (2007).
98. Tsai, A.C. *et al.* Late-onset combined homocystinuria and methylmalonic aciduria (*cblC*) and neuropsychiatric disturbance. *American journal of medical genetics. Part A* **143A**, 2430-4 (2007).
99. Bodamer, O.A., Rosenblatt, D.S., Appel, S.H. & Beaudet, A.L. Adult-onset combined methylmalonic aciduria and homocystinuria (*cblC*). *Neurology* **56**, 1113 (2001).
100. Komhoff, M. *et al.* Combined pulmonary hypertension and renal thrombotic microangiopathy in cobalamin C deficiency. *Pediatrics* **132**, e540-4 (2013).
101. Van Hove, J.L. *et al.* Cobalamin disorder Cbl-C presenting with late-onset thrombotic microangiopathy. *American journal of medical genetics* **111**, 195-201 (2002).
102. Carmel, R., Bedros, A.A., Mace, J.W. & Goodman, S.I. Congenital methylmalonic aciduria--homocystinuria with megaloblastic anemia: observations on response to hydroxocobalamin and on the effect of homocysteine and methionine on the deoxyuridine suppression test. *Blood* **55**, 570-9 (1980).
103. Roze, E. *et al.* Neuropsychiatric disturbances in presumed late-onset cobalamin C disease. *Archives of neurology* **60**, 1457-62 (2003).
104. Carrillo-Carrasco, N. & Venditti, C. Combined methylmalonic acidemia and homocystinuria, *cblC* type. II. Complications, pathophysiology, and outcomes. *Journal of Inherited Metabolic Disease* **35**, 103-114 (2012).

105. Froese, D.S., Zhang, J., Healy, S. & Gravel, R.A. Mechanism of vitamin B_{12}-responsiveness in *cblC* methylmalonic aciduria with homocystinuria. *Molecular genetics and metabolism* **98**, 338-43 (2009).
106. Richard, E. *et al.* Genetic and cellular studies of oxidative stress in methylmalonic aciduria (MMA) cobalamin deficiency type C (*cblC*) with homocystinuria (MMACHC). *Human Mutation* **30**, 1558-66 (2009).
107. Froese, D.S. *et al.* Thermolability of mutant MMACHC protein in the vitamin B_{12}-responsive *cblC* disorder. *Molecular genetics and metabolism* **100**, 29-36 (2010).
108. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biology* **9**, 1-21 (2011).
109. Morel, C.F., Lerner-Ellis, J.P. & Rosenblatt, D.S. Combined methylmalonic aciduria and homocystinuria (*cblC*): Phenotype genotype correlations and ethnic-specific observations. *Molecular Genetics and Metabolism* **88**, 315-321 (2006).
110. Lerner-Ellis, J.P. *et al.* Spectrum of mutations in *MMACHC*, allelic expression, and evidence for genotype-phenotype correlations. *Human Mutation* **30**, 1072-81 (2009).
111. Koutmos, M., Gherasim, C., Smith, J.L. & Banerjee, R. Structural Basis of Multifunctionality in a Vitamin B_{12}-processing Enzyme. *Journal of Biological Chemistry* **286**, 29780-29787 (2011).
112. Mah, W. *et al.* Subcellular location of MMACHC and MMADHC, two human proteins central to intracellular vitamin B_{12} metabolism. *Molecular genetics and metabolism* **108**, 112-8 (2013).
113. Plesa, M. *et al.* Interaction between MMACHC and MMADHC, two human proteins participating in intracellular vitamin B_{12} metabolism. *Molecular genetics and metabolism* **102**, 139-48 (2011).
114. Deme, J.C. *et al.* Structural features of recombinant MMADHC isoforms and their interactions with MMACHC, proteins of mammalian vitamin B_{12} metabolism. *Molecular genetics and metabolism* **107**, 352-62 (2012).
115. Fofou-Caillierez, M.B. *et al.* Interaction between methionine synthase isoforms and MMACHC: characterization in *cblG*-variant, *cblG* and *cblC* inherited causes of megaloblastic anaemia. *Human Molecular Genetics* **22**, 4591-601 (2013).
116. Bassford, P.J., Jr., Bradbeer, C., Kadner, R.J. & Schnaitman, C.A. Transport of vitamin B_{12} in tonB mutants of Escherichia coli. *Journal of bacteriology* **128**, 242-7 (1976).

117. Reynolds, P.R., Mottur, G.P. & Bradbeer, C. Transport of vitamin B_{12} in Escherichia coli. Some observations on the roles of the gene products of BtuC and TonB. *The Journal of biological chemistry* **255**, 4313-9 (1980).
118. Bell, P.E., Nau, C.D., Brown, J.T., Konisky, J. & Kadner, R.J. Genetic suppression demonstrates interaction of TonB protein with outer membrane transport proteins in Escherichia coli. *Journal of bacteriology* **172**, 3826-9 (1990).
119. Pupavac, M., Garcia, M.A., Rosenblatt, D.S. & Jerome-Majewska, L.A. Expression of *Mmachc* and *Mmadhc* during mouse organogenesis. *Molecular genetics and metabolism* **103**, 401-5 (2011).
120. Coulon, A., Chow, C.C., Singer, R.H. & Larson, D.R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nature reviews. Genetics* **14**, 572-84 (2013).
121. Shandilya, J. & Roberts, S.G. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochimica et biophysica acta* **1819**, 391-400 (2012).
122. Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics* **13**, 613-26 (2012).
123. Amir, R.E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methylCpG-binding protein 2. *Nature genetics* **23**, 185-188 (1999).
124. Samaco, R.C., Hogart, A. & LaSalle, J.M. Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of *UBE3A* and *GABRB3*. *Human Molecular Genetics* **14**, 483-92 (2005).
125. Kishi, N. & Macklis, J.D. *MECP2* is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Molecular and cellular neurosciences* **27**, 306-21 (2004).
126. Caballero, I.M. & Hendrich, B. MeCP2 in neurons: closing in on the causes of Rett syndrome. *Human Molecular Genetics* **14**, R19-R26 (2005).
127. Fajans, S.S., Bell, G.I. & Polonsky, K.S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *The New England journal of medicine* **345**, 971-80 (2001).
128. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (*MODY1*). *Nature* **384**, 458-60 (1996).
129. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (*MODY3*). *Nature* **384**, 455-8 (1996).

130. Stoffers, D.A., Ferrer, J., Clarke, W.L. & Habener, J.F. Early-onset type-II diabetes mellitus (MODY4) linked to *IPF1*. *Nature genetics* **17**, 138-9 (1997).
131. Horikawa, Y. *et al.* Mutation in hepatocyte nuclear factor-1 beta gene (*TCF2*) associated with MODY. *Nature genetics* **17**, 384-5 (1997).
132. Malecki, M.T. *et al.* Mutations in *NEUROD1* are associated with the development of type 2 diabetes mellitus. *Nature genetics* **23**, 323-8 (1999).
133. Neve, B. *et al.* Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4807-12 (2005).
134. Plengvidhya, N. *et al. PAX4* mutations in Thais with maturity onset diabetes of the young. *Journal of Clinical Endocrinology & Metabolism* **92**, 2821-2826 (2007).
135. Cerf, M.E. Transcription factors regulating beta-cell function. *European journal of endocrinology / European Federation of Endocrine Societies* **155**, 671-9 (2006).
136. Shih, D.Q. *et al.* Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism. *Diabetes* **50**, 2472-80 (2001).
137. Naya, F.J., Stellrecht, C.M. & Tsai, M.J. Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. *Genes & development* **9**, 1009-19 (1995).
138. Naya, F.J. *et al.* Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes & development* **11**, 2323-34 (1997).
139. Chae, J.H., Stein, G.H. & Lee, J.E. NeuroD: the predicted and the surprising. *Molecules and cells* **18**, 271-88 (2004).
140. Fuchs, T. *et al.* Mutations in the *THAP1* gene are responsible for DYT6 primary torsion dystonia. *Nature genetics* **41**, 286-288 (2009).
141. Djarmati, A. *et al.* Mutations in *THAP1* (DYT6) and generalised dystonia with prominent spasmodic dysphonia: a genetic screening study. *Lancet neurology* **8**, 447-52 (2009).
142. Abidi, F. *et al.* X-linked mental retardation with variable stature, head circumference, and testicular volume linked to Xq12-q21. *American Journal of Medical Genetics* **85**, 223-9 (1999).
143. Villard, L. *et al. XNP* mutation in a large family with Juberg-Marsidi syndrome. *Nature genetics* **12**, 359-60 (1996).
144. Stevenson, R.E., Abidi, F., Schwartz, C.E., Lubs, H.A. & Holmes, L.B. Holmes-Gang syndrome is allelic with XLMR-hypotonic face syndrome. *American Journal of Medical Genetics* **94**, 383-5 (2000).

145. Villard, L., Fontes, M., Ades, L.C. & Gecz, J. Identification of a mutation in the *XNP/ATR-X* gene in a family reported as Smith-Fineman-Myers syndrome. *American Journal of Medical Genetics* **91**, 83-5 (2000).
146. Abidi, F.E. *et al.* Mutation in the 5' alternatively spliced region of the *XNP/ATR-X* gene causes Chudley-Lowry syndrome. *European journal of human genetics : EJHG* **13**, 176-83 (2005).
147. Lossi, A.M. *et al.* Mutation of the *XNP/ATR-X* gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation: demonstration that the mutation is involved in the inactivation bias. *American journal of human genetics* **65**, 558-62 (1999).
148. Gibbons, R.J., Picketts, D.J., Villard, L. & Higgs, D.R. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* **80**, 837-45 (1995).
149. Gibbons, R.J. & Higgs, D.R. Molecular-clinical spectrum of the ATR-X syndrome. *American Journal of Medical Genetics* **97**, 204-12 (2000).
150. Gibbons, R. Alpha thalassaemia-mental retardation, X linked. *Orphanet journal of rare diseases* **1**, 15 (2006).
151. Gibbons, R.J. *et al.* Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS). *Nature genetics* **34**, 446-9 (2003).
152. Whitehouse, I. & Owen-Hughes, T. ATRX: Put me on repeat. *Cell* **143**, 335-6 (2010).
153. Huidobro, C., Fernandez, A.F. & Fraga, M.F. The role of genetics in the establishment and maintenance of the epigenome. *Cellular and molecular life sciences : CMLS* **70**, 1543-73 (2013).
154. Gibbons, R.J. *et al.* Mutations in *ATRX*, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nature genetics* **24**, 368-71 (2000).
155. Clynes, D. & Gibbons, R.J. ATRX and the replication of structured DNA. *Current opinion in genetics & development* (2013).
156. Wysocka, J. & Herr, W. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in biochemical sciences* **28**, 294-304 (2003).
157. Wilson, A.C., LaMarco, K., Peterson, M.G. & Herr, W. The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74**, 115-25 (1993).
158. Wilson, A.C., Peterson, M.G. & Herr, W. The HCF repeat is an unusual proteolytic cleavage signal. *Genes & development* **9**, 2445-58 (1995).

159. Michaud, J. *et al.* HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. *Genome research* **23**, 907-16 (2013).
160. Wysocka, J., Reilly, P.T. & Herr, W. Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Molecular and cellular biology* **21**, 3820-9 (2001).
161. Julien, E. & Herr, W. A switch in mitotic histone H4 lysine 20 methylation status is linked to M phase defects upon loss of HCF-1. *Molecular cell* **14**, 713-25 (2004).
162. Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508-12 (2010).
163. Dou, Y. *et al.* Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873-85 (2005).
164. Dejosez, M. *et al.* Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes & development* **24**, 1479-84 (2010).
165. Huang, L. *et al.* A noncoding, regulatory mutation implicates *HCFC1* in nonsyndromic intellectual disability. *American journal of human genetics* **91**, 694-702 (2012).
166. Glinsky, G.V., Berezovska, O. & Glinskii, A.B. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *The Journal of clinical investigation* **115**, 1503-21 (2005).
167. Zargar, Z. & Tyagi, S. Role of host cell factor-1 in cell cycle regulation. *Transcription* **3**, 187-92 (2012).
168. Tyagi, S., Chabes, A.L., Wysocka, J. & Herr, W. E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Molecular cell* **27**, 107-19 (2007).
169. Parker, J.B., Palchaudhuri, S., Yin, H., Wei, J. & Chakravarti, D. A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function. *Molecular and cellular biology* **32**, 1654-70 (2012).
170. Hammer, S.E., Strehl, S. & Hagemann, S. Homologs of Drosophila P Transposons Were Mobile in Zebrafish but Have Been Domesticated in a Common Ancestor of Chicken and Human. *Molecular Biology and Evolution* **22**, 833-844 (2005).
171. Dejosez, M. *et al.* Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* **133**, 1162-74 (2008).

172. Zhu, C.Y. *et al.* Cell growth suppression by thanatos-associated protein 11(THAP11) is mediated by transcriptional downregulation of c-Myc. *Cell death and differentiation* **16**, 395-405 (2009).
173. Zwaka, T.P. Ronin and caspases in embryonic stem cells: a new perspective on regulation of the pluripotent state. *Cold Spring Harbor symposia on quantitative biology* **73**, 163-9 (2008).
174. Gervais, V., Campagne, S., Durand, J., Muller, I. & Milon, A. NMR studies of a new family of DNA binding proteins: the THAP proteins. *Journal of biomolecular NMR* **56**, 3-15 (2013).
175. Yu, H.-C. *et al.* A Novel X-Linked Cobalamin Disorder Caused by Mutations in Transcriptional Coregulator *HCFC1*. (2013).
176. Kibbe, W. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35**(2007).
177. NCBI. BLAST. Vol. 2012 (2012).
178. Fujita, P. *et al.* The UCSC Genome Browser database: update 2011. *nucleic Acids Research* **39**, D876-D882 (2010).
179. Lohse, M. *et al.* RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic acids research* **40**, W622-7 (2012).
180. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
181. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25 (2009).
182. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-5 (2010).
183. Anders, S. HTSeq: Analysing high-throughput sequencing data with Python. (2010).
184. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106 (2010).
185. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
186. Gizicki, R. *et al.* Long-term visual outcome of methylmalonic aciduria and homocystinuria, cobalamin C type. *Ophthalmology* **121**, 381-6 (2014).
187. Morel, C.F., Lerner-Ellis, J.P. & Rosenblatt, D.S. Combined methylmalonic aciduria and homocystinuria (*cblC*): phenotype-genotype correlations and ethnic-specific observations. *Molecular genetics and metabolism* **88**, 315-21 (2006).

188. Sengel, C., Gavarini, S., Sharma, N., Ozelius, L.J. & Bragg, D.C. Dimerization of the DYT6 dystonia protein, THAP1, requires residues within the coiled-coil domain. *Journal of neurochemistry* **118**, 1087-100 (2011).
189. Di Liegro, C.M., Schiera, G. & Di Liegro, I. Regulation of mRNA transport, localization and translation in the nervous system of mammals (Review). *Int J Mol Med* **33**, 747-62 (2014).
190. dos Remedios, C.G. *et al.* Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* **83**, 433-73 (2003).
191. Watakabe, A., Kobayashi, R. & Helfman, D.M. N-tropomodulin: a novel isoform of tropomodulin identified as the major binding protein to brain tropomyosin. *Journal of cell science* **109 (Pt 9)**, 2299-310 (1996).
192. Moroz, N., Guillaud, L., Desai, B. & Kostyukova, A.S. Mutations changing tropomodulin affinity for tropomyosin alter neurite formation and extension. *PeerJ* **1**, e7 (2013).
193. Coste, B. *et al.* Gain-of-function mutations in the mechanically activated ion channel PIEZO2 cause a subtype of Distal Arthrogryposis. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 4667-72 (2013).
194. Sim, E., Abuhammad, A. & Ryan, A. Arylamine N-acetyltransferases: From Drug Metabolism and Pharmacogenetics to Drug Discovery. *British journal of pharmacology* (2014).
195. Minchin, R.F. Acetylation of p-aminobenzoylglutamate, a folic acid catabolite, by recombinant human arylamine N-acetyltransferase and U937 cells. *The Biochemical journal* **307 (Pt 1)**, 1-3 (1995).

APPENDIX A
PCR Conditions and Primer Sequences

	MMACHC						
	Exon 1	Exon 2	Exon 3	Exon 4A	Exon 4B	Exon 5A	cDNA
10X Buffer	5	5	5	5	5	5	5
$MgCl_2$	4	0	4	4	4	2	2
Water	36	35	34	36	36	37	38
Taq*	0.2 HS	0.2 HS	0.2 HS	0.2 HS	0.2 HS	0.2 HS	0.2HS
dNTPs	1	1	1	1	1	1	1
Primers	1	1	1	1	1	1	1
Q Buffer	0	5	0	0	0	0	0
Annealing T (°C)	58	58	58	58	58	65	61
Size (bp)	680	467	395	694	395	685	621

	HCFC1			*THAP11*	
	Exon 1	Exon 2	Exon 3	Exon 1A	Exon 1B
10X Buffer	5	5	5	5	5
Mg	0	4	4	0	0
Water	40	36	36	35	35
Taq*	0.2 HS	0.2 HS	0.2 T	0.2 HS	0.2 HS
dNTPs	1	1	1	1	1
Primers	1	1	1	1	1
Q Buffer	0	0	0	5	5
Annealing T (°C)	65	58	65	58	58
Size (bp)	726	547	471	863	723

All measurements are in microliters (µL)

*HS=HotStar Taq Polymerase, T=Taq Polymerase

Standard program:

Step		Taq	HotStar
1	Initial	94°C 3min	96°C 15min
2	Denature	94°C 30sec	96°C 30sec
3	Anneal	X°C 1min	X°C 1min
4	Extend	72°C 1min	72°C 1min
5	Repeat 2-4	40 cycles	40 cycles
6	Final Extension	72°C 10min	72°C 10min

Primer Sequences 5'→3'

MMACHC 1F	GGGATACCGTGATGATACGC
MMACHC 1R	GAACCCAGGAGGATCAGAGG
MMACHC 2F	TGCATCACATAGCGTCAGTG
MMACHC 2R	AGCCTGGCTTTAGGGTATCA
MMACHC 3F	TCATGTTTTCCCTTCTGAGGA
MMACHC 3R	CAAAGCTAATTTGTTCTGGGTTG
MMACHC 4AF	AGGCCTAGCTTGCAATGATG
MMACHC 4AR	GAAGGCAGATGGGAATTCTG
MMACHC 4BF	TTTGGCAAAGCAAAAGGTT
MMACHC 4BR	CAAGATGGGTGGATCACGA
MMACHC 5AF	AGCCTGGCCAATACAGTGAA
MMACHC 5AR	AGCCTTCCCTTGGTTCTAGC
MMACHC cDNA F	CGCTATGTCCTTTTGGCTTC
MMACHC cDNA R	AGGCCTTCTGCTCTTCTGAGT
HCFC1 1F	GGGAAGAGCCAAAGGGAAGG
HCFC1 1R	GTCGGACTAGGGGCTTTCTT
HCFC1 2F	TTGCACAGTAACTTCCCGCT
HCFC1 2R	CTGCCTTCTTCTGAGGTGGG
HCFC1 3F	CACCTCAGAAGAAGGCAGCA
HCFC1 3R	CGGTGCCGCTCTCTCATTTA
THAP11 1AF	CTCTTCGGAGGCTATCGCAG
THAP11 1AR	CTACAGTGGCCTGAAGGGTG
THAP11 1BF	CACCCTCTGCCTCCACTG
THAP11 1AR	CCAGCTTCCACTTCAGGAGG

APPENDIX B
Publications, Presentations and Abstracts

Publication H-C Yu, JL Sloan, G Scharer, A Brebner, A Quintana, NP Achilly, I Manoli, CR Coughlin II, EA Geiger, U Schneck, D Watkins, JL VanHove, B Fowler, MR Baumgartner, DS Rosenblatt, CP Venditti, TH Shaikh. An X-linked cobalamin disorder caused by mutations in transcriptional coregulator *HCFC1*. **American Journal of Human Genetics** (2013), 93(3): 506-514.

Publication ML Illson, L Dempsey-Nunez, J Kent, Q Huang, A Brebner, ML Raff, D Watkins, BM Gilfix, CT Wittwer, DS Rosenblatt. High resolution melting analysis of the *MMAB* gene in *cblB* patients and in those with undiagnosed methylmalonic aciduria. **Molecular Genetics and Metabolism** (2013), 110(1-2): 86-89.

Publication CM Armour, A Brebner, D Watkins, MT Geraghty, A Chan, DS Rosenblatt. Good clinical outcome in an atypical *cblF* patient identified by expanded newborn screening. **Pediatrics** (2013), doi: 10.1542/peds.2013-0105.

Publication L Dempsey-Nunez, ML Illson, J Kent, Q Huang, A Brebner, D Watkins, BM Gilfix, CT Wittwer, DS Rosenblatt. High resolution melting analysis of the *MMAA* gene in patients with *cblA* and in those with undiagnosed methylmalonic aciduria. **Molecular Genetics and Metabolism** (2012), 107(3): 363-367.

Presentation A Brebner, D Watkins, H-C Yu, JL S, A Quintana, NP Achilly, EA Geiger, CP Venditti, TH Shaikh, DS Rosenblatt. Mutations in the transcription factor *HCFC1* in patients with the *cblX* inborn error of vitamin B_{12} metabolism. **Human Genetics Research Day**, June 6 2013, Montreal QC.

Presentation A Brebner, D Watkins, H-C Yu, JL S, A Quintana, NP Achilly, EA Geiger, CP Venditti, TH Shaikh, DS Rosenblatt. Mutations in the transcription factor *HCFC1* in patients with the *cblX* inborn error of vitamin B_{12} metabolism. **Garrod Symposium**, May 30 2013, Sherbrooke QC.

Presentation	A Brebner, C Ficicioglu, D Watkins, T Pastinen, DS Rosenblatt. Mutations in an atypical patient with combined homocystinuria and methylmalonic aciduria (*cblC*). **Biology Honours Research Symposium**, April 30 2012, Montreal QC
Published Abstract Poster	A Brebner, D Watkins, DS Rosenblatt. Molecular analysis of patients diagnosed as *cblC*. **63rd Annual American Society of Human Genetics Meeting**, 22-26 October 2013, Boston MA. [Published Online]
Published Abstract	TH Shaikh, H-C Yu, JL Sloan, G Scharer, A Brebner, A Quintana, NP Achilly, I Manoli, CR Coughlin, EA Geiger, U Schneck, D Watkins, JL Van Hove, B Fowler, MR Baumgartner, DS Rosenblatt, CP Venditti. Mutations in *HCFC1* a transcriptional coregulator causes a novel X-linked cobalamin disorder (*cblX*) with a severe neurological phenotype. **63rd Annual American Society of Human Genetics Meeting**, 22-26 October 2013, Boston MA. [Published Online]
Published Abstract Poster	A Brebner, D Watkins, DS Rosenblatt. Novel mutations in patients with the *cblC* inborn error of cobalamin metabolism. **12th International Congress of Inborn Errors of Metabolism**, 3-6 September 2013, Barcelona Spain. Published in Journal of Inherited Metabolic Disease (2013), 36(2): S258
Published Abstract	JL Sloan, H-C Yu, G Scharer, A Brebner, A Quintana, NP Achilly, I Manoli, CR Coughlin, EA Geiger, U Schneck, D Watkins, JL Van Hove, B Fowler, MR Baumgartner, DS Rosenblatt, CP Venditti, TH Shaikh. Mutations in *HCFC1*, a transcriptional co-regulator, causes a novel X-linked cobalamin disorder (*cblX*) with a severe neurological phenotype. **12th International Congress of Inborn Errors of Metabolism**, 3-6 September, Barcelona Spain. Published in Journal of Inherited Metabolic Disease (2013), 36(2): S105
Published Abstract Poster	A Brebner, C Ficicioglu, D Watkins, T Pastinen, DS Rosenblatt. Somatic cell and molecular studies on an atypical patient with combined homocystinuria and methylmalonic aciduria (*cblC*). **62nd Annual American Society of Human Genetics Meeting**, 6-10 November 2012, San Francisco CA. [Published Online]

Poster Presentation	A Brebner, D Watkins, H-C Yu, JL Sloan, A Quintana, NP Achilly, EA Geiger, CP Venditti, TH Shaikh, DS Rosenblatt. Mutations in the transcription factor *HCFC1* in patients with the *cblX* inborn error of vitamin B_{12} metabolism. **2nd Annual Canadian Human and Statistical Genetics Meeting**, 21-24 April 2013, Esterel QC. Masters project
Poster Presentation	A Brebner, D Watkins, C Ficicioglu, D Watkins, T Pastinen, DS Rosenblatt. Search for non-coding mutations in an atypical patient with combined homocystinuria and methylmalonic aciduria (*cblC*). **Réseau de médecine génétique appliquée**, 22-24 May 2012, Montreal QC. Honours project

APPENDIX C
Article Yu *et al.* 2013

Thesis Author's Contribution: Collected patients for mutation analysis, characterized clinical and biochemical phenotype, edited manuscript

ARTICLE

An X-Linked Cobalamin Disorder Caused by Mutations in Transcriptional Coregulator *HCFC1*

Hung-Chun Yu,[1] Jennifer L. Sloan,[2] Gunter Scharer,[1,3,4,9] Alison Brebner,[5] Anita M. Quintana,[1] Nathan P. Achilly,[2] Irini Manoli,[2] Curtis R. Coughlin II,[1,3] Elizabeth A. Geiger,[1] Una Schneck,[1] David Watkins,[5] Terttu Suormala,[6,7] Johan L.K. Van Hove,[1,3] Brian Fowler,[6,7] Matthias R. Baumgartner,[6,7,8] David S. Rosenblatt,[5] Charles P. Venditti,[2] and Tamim H. Shaikh[1,3,4,*]

Derivatives of vitamin B_{12} (cobalamin) are essential cofactors for enzymes required in intermediary metabolism. Defects in cobalamin metabolism lead to disorders characterized by the accumulation of methylmalonic acid and/or homocysteine in blood and urine. The most common inborn error of cobalamin metabolism, combined methylmalonic acidemia and hyperhomocysteinemia, *cblC* type, is caused by mutations in *MMACHC*. However, several individuals with presumed *cblC* based on cellular and biochemical analysis do not have mutations in *MMACHC*. We used exome sequencing to identify the genetic basis of an X-linked form of combined methylmalonic acidemia and hyperhomocysteinemia, designated *cblX*. A missense mutation in a global transcriptional coregulator, *HCFC1*, was identified in the index case. Additional male subjects were ascertained through two international diagnostic laboratories, and 13/17 had one of five distinct missense mutations affecting three highly conserved amino acids within the HCFC1 kelch domain. A common phenotype of severe neurological symptoms including intractable epilepsy and profound neurocognitive impairment, along with variable biochemical manifestations, was observed in all affected subjects compared to individuals with early-onset *cblC*. The severe reduction in *MMACHC* mRNA and protein within subject fibroblast lines suggested a role for HCFC1 in transcriptional regulation of *MMACHC*, which was further supported by the identification of consensus HCFC1 binding sites in *MMACHC*. Furthermore, siRNA-mediated knockdown of *HCFC1* expression resulted in the coordinate downregulation of *MMACHC* mRNA. This X-linked disorder demonstrates a distinct disease mechanism by which transcriptional dysregulation leads to an inborn error of metabolism with a complex clinical phenotype.

Introduction

Cobalamin-derived cofactors, 5′-adenosylcobalamin (Ado-Cbl) and methylcobalamin (Me-Cbl), are essential for the activity of methylmalonyl-CoA mutase (MUT [ENZYME EC 5.4.99.2]) and methionine synthase (MTR [ENZYME EC 2.1.1.13]), respectively. MUT functions in the catabolism of branched-chain amino acids and odd-chain fatty acids into the Krebs cycle, and MTR catalyzes the remethylation of homocysteine to methionine. Mutations in *MUT* (MIM 609058) or *MTR* (MIM 156570) or defects in either the intracellular transport of cobalamin or the synthesis of the active enzymatic cofactors result in one of nine distinct inborn errors of metabolism, historically designated by the cellular complementation class (*mut*, *cblA*–*cblG*, or *cblJ*).[1] The disorders that affect both the synthesis and the transport of 5′-Ado-Cbl and Me-Cbl (*cblC* [MIM 277400]), *cblD* combined [MIM 277410], *cblF* [MIM 277380], and *cblJ* [MIM 614857]) share the characteristic biochemical findings of combined methylmalonic acidemia and hyperhomocysteinemia.

Combined methylmalonic acidemia and hyperhomocysteinemia, *cblC* type, one of the first recognized and most common inborn errors of cobalamin metabolism, is caused by mutations in *MMACHC*[2] (MIM 609831). The clinical phenotype of *cblC* is diverse and can feature neurologic, renal, cardiac, hematologic, and ophthalmologic manifestations.[3–5] Although the clinical phenotypes of *cblD* combined, *cblF*, and *cblJ* defects are not as well delineated, most reported cases have been phenotypically similar to *cblC*.[6–8] The genes associated with these disorders have been discovered.[6–9] However, several individuals who have a cellular, but not molecular, diagnosis of *cblC* still exist,[2,9] suggesting genetic heterogeneity within this complementation group. *MMACHC* sequence analysis did not identify causative mutations in 14/204[9] or 3/118[2] reported individuals with biochemically confirmed *cblC* deficiency.

Here, we describe the identification of an X-linked form of combined methylmalonic acidemia and hyperhomocysteinemia, designated *cblX*. Exome sequencing of a male subject initially diagnosed with *cblC* by complementation analysis but who had no mutations in *MMACHC* identified a mutation in *HCFC1* (MIM 300019), a coregulator of the zinc-finger transcription factor THAP11 (also known as RONIN).[10,11] Sanger sequencing of *HFCF1* in 17 additional male subjects revealed that 13 harbored one of five pathogenic missense *HCFC1* mutations. Functional analysis

[1]Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO 80045, USA; [2]Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [3]Section of Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA; [4]Colorado Intellectual and Developmental Disabilities Research Center, University of Colorado School of Medicine, Aurora, CO 80045, USA; [5]Department of Human Genetics, McGill University, Montreal, Quebec H3A 1B1, Canada; [6]Division of Metabolism, University Children's Hospital, Zürich 8032, Switzerland; [7]Children's Research Center, University Children's Hospital, Zürich 8032, Switzerland; [8]Zürich Center for Integrative Human Physiology, University of Zürich, Zürich 8057, Switzerland
[9]Present address: Medical College of Wisconsin, Department of Pediatrics, Section of Clinical Genetics, Milwaukee, WI 53226, USA
*Correspondence: tamim.shaikh@ucdenver.edu
http://dx.doi.org/10.1016/j.ajhg.2013.07.022.

further implicated HCFC1 in transcriptional regulation of *MMACHC*.

Subjects and Methods

Human Subjects

All human subject samples used in this study were collected after informed consent was obtained. Subject 1 and his parents were enrolled into a research protocol (COMIRB no. 07-0386) approved by the institutional review board at the University of Colorado, School of Medicine. Subject 1, subject 11, and their parents were enrolled in National Institutes of Health study "Clinical and Basic Investigations of Methylmalonic Acidemia and Related Disorders" (clinicaltrials.gov identifier NCT00078078) in compliance with the Helsinki Declaration and provided informed consent. The studies at McGill University Health Centre (Montreal) and University Children's Hospital (Zurich) were approved by their respective ethics committees. Genomic DNA was obtained from either peripheral-blood lymphocytes or cultured skin fibroblasts.

Whole-Exome Sequencing

Whole-exome sequencing (WES) on subject 1 and his parents was performed with Nimblegen SeqCap EZ Exome v.2.0 and was followed by sequencing on an Illumina HiSeq 2000. Approximately 50 million 90 bp paired-end reads (>50×) were obtained. Sequence reads were first mapped to the human reference genome (hg19 assembly, UCSC Genome Browser) with the Burrows-Wheeler Aligner[12] and then visualized with the Integrative Genomics Viewer.[13] The utilities in SAMtools[14] were used for variant calling, and variants were annotated with SeattleSeq. Further filtering and testing of the inheritance model were done with tools in Galaxy.[15] Single-nucleotide variants (SNVs) were filtered for retaining calls that met the following criteria: SNP and consensus scores >50, read coverage >8, and >25% of the reads containing the variant call. Bases with a PHRED-scaled score <20 were removed from the analysis. For homozygous and hemizygous variant calls, >80% reads were required to contain the variant, whereas for heterozygous variant calls, the number of reads containing the variant call ranged between 25% and 80%. Indels < 50 bp were filtered on the basis of similar criteria, except that the SNP and consensus scores were required to be >100. Variants found in segmental duplications or simple- or low-complexity repeats were removed because of the higher likelihood of mapping errors. Sequence data from parental samples were used as an additional filter for confirmation of variant calls in subject 1. The filtering criteria for variant calling in parental data were less stringent than those in subject 1 so that erroneous classification of variants as unique to subject 1 could be minimized. Thus, the criteria for parental data included SNP score > 5 for SNVs and SNP score > 10 for indels and required that at least two reads contain the variant call. Variants were filtered against dbSNP build 135 and 1000 Genomes (November 23, 2010 release). The sequence data from the family was then used for testing for causal variants under different inheritance models, including a de novo mutation in a dominant model and compound-heterozygous, homozygous, and X-linked hemizygous mutations in recessive models.

Sanger Sequencing

Variants identified in *HCFC1* and *TTN* (MIM 188840) were further validated by Sanger sequencing in subject 1 and his parents. Mutations detected in *HCFC1* and *TTN* were named with the use of RefSeq cDNA accession numbers NM_005334.2 and NM_133378.4, respectively. Primers were designed to amplify and sequence the 26 coding exons of *HCFC1* in the 17 additional subjects and available parents (Table S1, available online). Genomic DNA (100 ng) was amplified with the following PCR conditions: Promega GoTaq Hot Start kit with 1× Master Mix and 400 nM of each primer. PCR began with an initial cycle at 95°C for 3 min and 30 subsequent cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min and finished with an extension at 72°C for 5 min. Amplified PCR products were sequenced with the PCR primers as sequencing primers on an ABI PRISM 3730xl by a commercial sequencing facility.

Protein Alignment

HCFC1 sequences in human (RefSeq NP_005325.2), chimpanzee (RefSeq XP_521330.3), mouse (RefSeq NP_032250.2), rat (RefSeq NP_001132979.1), frog (RefSeq XP_002937269.1), and two zebrafish paralogs (RefSeq NP_001122009.1 and NP_001038529.1) were aligned with ClustalW.[16]

Protein Modeling

Predicted models of the wild-type and altered HCFC1 (UniProtKB P51610) kelch domain (amino acids 42–342) were created by Modeller[17] with human KLHL12 (Protein Data Bank ID 2VPJ) as a modeling template. KLHL12 was chosen by CPHmodels[18] because it had an available protein model with the highest homology with the HCFC1 kelch domain. The resulting structures were visualized with Chimera.[19]

HCFC1 and THAP11 Binding Sites

Binding regions of HCFC1 and THAP11 were obtained from chromatin immunoprecipitation sequencing (ChIP-seq) data in mouse embryonic stem cells.[10] Mouse genomic coordinates from the mm8 assembly of the UCSC Genome Browser were converted to the orthologous human hg19 assembly with the LiftOver tool from the UCSC Genome Browser.[20] A highly conserved 15 bp (5′-CTGGGARWTGTAGTY-3′) THAP11 binding motif[10] was identified by fuzznuc[21] in the human genome. A maximum of 2 bp mismatch was allowed.

Analysis of RNA Expression

For quantitative evaluation of gene expression, total RNA was isolated from human fibroblasts and reverse transcribed. The two control samples used were from healthy individuals with no known biochemical or neurological phenotypes. The Promega GoScript reverse-transcription system was used for converting RNA to cDNA with random primers. Assays for quantitative PCR (qPCR) were designed with the Roche Universal ProbeLibrary Design Center, and Roche Universal Probes and FastStart Universal Probe Master Mix with Rox were used. The Roche Universal ProbeLibrary Human ACTB (β-actin) Gene Assay was used as an endogenous control (catalog no. 05046165001). Primers (Table S1) were synthesized by Integrated DNA Technologies. Primer pairs that showed an amplification efficiency within ± 10% of each other and a coefficient of correlation between 0.95 and 1.0 were used for quantification. ABI 7500 Software version 2.0.5 was used to calculate the relative expression levels with the $\Delta\Delta C_T$ method. The reactions were all carried out in technical triplicates. The experiment was also repeated in three biological replicates. Error bars represent the SEM of relative expression

levels. Statistical significance was calculated with a two-tailed t test.

Immunoblot Analysis

Control human dermal fibroblasts (C-013-5C) were obtained from Life Technologies. Human fibroblasts were grown to ~95% confluence in 10 cm plates and lysed in T-PER containing a protease-inhibitor cocktail (Thermo Fisher Scientific). Crude lysates were clarified by centrifugation at 13,000 rpm for 10 min at 4°C. Seventy-five micrograms of lysate was analyzed by immunoblots with a rabbit polyclonal MMACHC (ab96195; Abcam) antibody at a dilution of 1:1,000 or rabbit polyclonal HCFC1 antibody (A301-400A; Bethyl Laboratories) at a dilution of 1:2,000. The blot was stripped with Restore Western Blot Stripping Buffer (Thermo Fisher Scientific) and reprobed with a rabbit polyclonal β-actin antibody at a dilution of 1:3,000 as a loading control. Secondary antibody was visualized by chemiluminescence detection (Thermo Fisher Scientific). Immunoblots for *HCFC1* siRNA knockdown experiments were performed on cell lysates obtained from human embryonic kidney 293 (HEK293) cells (American Type Culture Collection) treated with either *HCFC1* siRNA or scrambled siRNA in three biological replicates. Cells were harvested for immunoblotting 4 days after transfection. The antibodies used for immunoblot analysis of HCFC1 and β-actin were the same as above.

HCFC1 siRNA Knockdown

HEK293 cells were transfected with Trilencer-27 siRNA duplexes (Origene) with the use of the siTran transfection reagent and incubated for 48 hr. RNA was isolated from *HCFC1* siRNA-treated cells and scrambled-siRNA-treated control cells with the RNeasy Micro Kit (QIAGEN). Total RNA was quantitated in each individual sample, and 1 μg of total RNA was used for producing cDNA with random hexamer primers with the GoScript Reverse Transcription System (Promega) according to the manufacturer's instructions. Quantitative PCR was performed as described above. The reactions were all carried out in technical triplicates. The experiment was also repeated in three biological replicates. Error bars represent the SEM of relative expression levels. Statistical significance was calculated with a two-tailed t test.

MMA Production by Cultured *cblX* Fibroblasts

A modified chemical-stimulation study was performed as described previously.[22] Six-well tissue-culture plates were seeded at a density of 2×10^5 or 5×10^5 cells per well in high glucose (4 g/l) DMEM supplemented with 10% fetal bovine serum, penicillin streptomycin, L-glutamine, and sodium pyruvate. The next day, the DMEM growth medium was removed and replaced with 1 ml of DMEM growth medium containing sodium propionate (10 mM) or sodium propionate (10 mM) plus hydroxocobalamin (1 μg/ml). After 72 hr, the medium was collected for gas chromatography–mass spectrometry analysis of MMA and measured in triplicate. Cell viability was determined by visual inspection and noted to be equal in all groups.

Results

Clinical Data

The clinical features of subjects are shown in Table 1. A common phenotype of severe neurological symptoms, which primarily include intractable epilepsy and neurocognitive impairment, was observed in all affected males (Table 1). Disease onset was either in the prenatal period (n = 3) or in early infancy (n = 11), and severe neurological manifestations were present in all subjects. Seizures (n = 12), profound developmental delay (n = 11), microcephaly (n = 7), movement disorders (n = 4), and a history of maternal male infant deaths (n = 3) were reported. Biochemical perturbations included hyperhomocysteinemia, methylmalonic acidemia, and increased propionylcarnitine (Table 1 and Table S2). Laboratory values were more variable than seen in individuals with early-onset *cblC*[4] (Figure 1) given that five *cblX* subjects had normal plasma total homocysteine. Cellular biochemical studies were performed in all subjects and revealed diminished 1-[^{14}C]-propionate and 5-[^{14}C]-methyl-THF incorporation (which improved with hydroxocobalamin supplementation), impaired synthesis of Ado-Cbl and Me-Cbl (Table S2), and failure to complement respective *cblC* cell lines.

WES

We selected one male subject (subject 1) with presumed *cblC* for WES analysis because of his atypical neurological phenotype, which included a gyral cortical malformation that required surgical resection. We first filtered out common variants present in dbSNP and 1000 Genomes and focused on nonsynonymous coding variants, coding indels, and variants affecting splice sites, resulting in 161 rare variants that were considered for further analysis (Table S3). Parental WES data were used for identifying possible pathogenic variants under various inheritance models, including dominant (de novo mutations) and recessive (compound-heterozygous, homozygous, and X-linked hemizygous mutations) models.

Under the recessive model, we identified two candidate genes with potentially pathogenic mutations. The first, *TTN*, encodes Titin, a giant protein involved in the assembly of cardiac and skeletal muscle. Subject 1 had compound-heterozygous missense changes in *TTN*: c.61634G>A (p.Arg20545Gln) inherited from his mother and c.77236G>A (p.Val25746Ile) inherited from his father. The only known function of *TTN* in muscle assembly made it unlikely to play a role in disease etiology. The second candidate, *HCFC1* on chromosome X, had a nonsynonymous, missense variant in exon 3, c.344C>T (p.Ala115Val). Sanger sequencing validated the variant in subject 1 and his unaffected carrier mother. This mutation was not found in the subject's maternal grandparents, suggesting that the mutation might have originated de novo in his mother (Figure S1A). The pleiotropic function of HCFC1 coupled with the recent description of an X-linked form of mental retardation (MRX3 [MIM 309541]) caused by its altered expression prompted further investigations.[23]

HCFC1 Mutation Screening

Seventeen additional males, the majority of whom had been diagnosed with *cblC* deficiency by cellular studies

Table 1. Clinical and Biochemical Features of Male Subjects with *HCFC1* Variants

Subject	Age of Onset	Clinical Symptoms	Genetic Testing (Negative)	*HCFC1* Variant	Plasma tHcy[a] (μmol/l)	Serum MMA[b] (μmol/l)	Urine MMA[c] (mmol/mol creatinine)
1	4 months	severe developmental delay, infantile spasms with hypsarrythmia, gyral cortical malformation, microcephaly, chorea, undescended testes, megacolon	*MMACHC*	c.344C>T[d] (p.Ala115Val)	141	21.5	ND
2	4 months	severe developmental delay, intractable epilepsy, choreoathetosis, microcephaly, FTT	*MMACHC*	c.344C>T (p.Ala115Val)	ND	ND	650
3	prenatal	severe developmental delay, neonatal epilepsy, choreoathetosis, congenital microcephaly, FTT	*MMACHC*	c.344C>T (p.Ala115Val)	ND	ND	elevated
4	<3 months	infantile spasms with hypsarrythmia, absent development	*MMACHC*	c.344C>T (p.Ala115Val)	5	ND	201
5	2 months	severe developmental delay, epilepsy, choreoathetosis, neutropenia	*MMACHC*	c.344C>T (p.Ala115Val)	WNL	ND	626
6	prenatal	severe developmental delay, infantile spasms with hypsarrythmia, microcephaly, IUGR	*MMACHC*	c.344C>T (p.Ala115Val)	ND	22.7	192
7	4 months	epilepsy, hypotonia	*MMACHC*	c.344C>T (p.Ala115Val)	118	11.6	ND
8	<3 months	severe developmental delay, intractable epilepsy, hypospadias	*MMACHC*	c.344C>T (p.Ala115Val)	80	ND	elevated
9	<3 months	severe developmental delay, epilepsy, no eye contact, muscular hypotonia, microcephaly	*MMACHC*	c.344C>T (p.Ala115Val)	61	5.5	ND
10	2 weeks	acute neurologic and metabolic decompensation, ketoacidosis with hyperammonemia, microcephaly, abnormal EEG, death in infancy	*MMACHC*	c.343G>A (p.Ala115Thr)	ND	elevated	elevated
11	5 months	severe developmental delay, intractable epilepsy, choreoathetosis, FTT, ballismus	*MMADHC*, *MMACHC*	c.218C>T[e] (p.Ala73Val)	9	5.7	127
12	9 weeks	severe developmental delay, epilepsy, muscular hypotonia, generalized disturbance and delay of myelination on MRI at age 8 years	*MMACHC*	c.218C>T (p.Ala73Val)	13	ND	329
13	5 weeks	severe developmental delay, epilepsy, FTT	*MMACHC*	c.217G>A (p.Ala73Thr)	WNL	ND	elevated
14	prenatal	seizures, blindness, hearing impairment, slight dysmorphia, congenital microcephaly, IUGR, FTT	*MMACHC*	c.202C>G (p.Gln68Glu)	87	23.0	elevated

Abbreviations are as follows: FTT, failure to thrive; IUGR, intrauterine growth retardation; ND, not determined; and WNL, within normal limits.
[a]Normal is <13 μmol/l.
[b]Normal is <0.4 μmol/l.
[c]Normal is undetectable.
[d]Mother was found to carry a de novo variant.
[e]De novo variant.

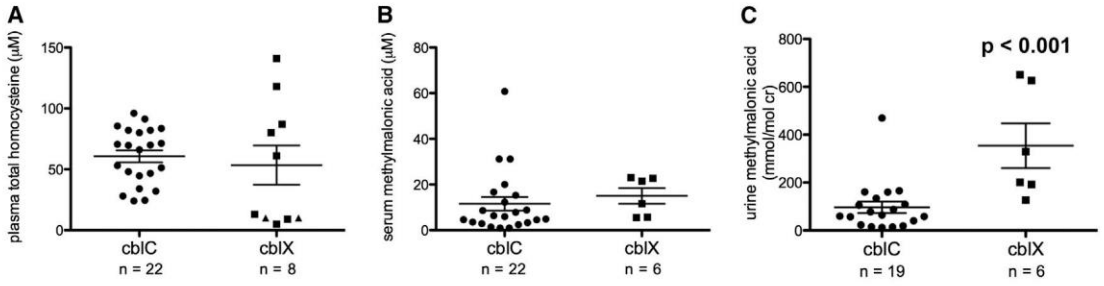

Figure 1. Metabolite Measurements in *cblX* Subjects Compared to Treated, Early-Onset *cblC* Subjects
Plasma total homocysteine (tHcy) and serum and urine methylmalonic acid (MMA) were measured in *cblX* subjects (single measurements) and 23 subjects with early-onset *cblC* deficiency (1–13 measurements per subject). In cases where multiple measurements were available for the same subject, the mean of all readings was used. n = the total number of subjects in each group.
(A) There was no statically significant difference in plasma tHcy measurements between *cblC* subjects and *cblX* subjects (*cblC*, 60.63 ± 4.88 μM; *cblX*, 64.25 ± 18.30 μM [mean ± SEM]). Normal levels of plasma tHcy are <13 μM. Triangles indicate two subjects with reportedly normal tHcy (no concentration was provided to the referring diagnostic laboratory).
(B) There was no significant difference in serum MMA measurements between *cblC* subjects and *cblX* subjects (*cblC*, 11.57 ± 2.99 μM; *cblX*, 15.00 ± 3.43 μM [mean ± SEM]). Normal levels of serum MMA are <0.4 μM.
(C) *cblX* subjects had higher urine MMA (t test $p < 0.001$) than did *cblC* subjects (*cblC*, 96.52 ± 23.89 mmol/mol creatinine; *cblX*, 354.20 ± 93.70 mmol/mol creatinine [mean ± SEM]). Normal levels of urine MMA are <4 mmol/mol creatinine.

but who likewise did not have mutations in *MMACHC* (Table 1), were identified through a retrospective review of laboratory and clinical data from two international referral centers. Sanger sequencing of the coding exons and splice sites of *HCFC1* identified one of five distinct pathogenic missense mutations in 13/17 subjects: three in exon 2 (c.202C>G [p.Gln68Glu], c.217G>A [p.Ala73Thr], and c.218C>T [p.Ala73Val]) and two in exon 3 (c.343G>A [p.Ala115Thr] and c.344C>T [p.Ala115Val]). The missense changes affect three highly conserved amino acids: Gln68, Ala73, and Ala115 (Figure 2). The mutation present in the index case, c.344C>T (p.Ala115Val), was the most frequent and was present in nine of the subjects.

Parental samples were only available for subject 1, whose mother was a carrier, and subject 11, whose parents were both negative for the c.218C>T (p.Ala73Val) change (the fact that subject 11's parents were negative for the mutation suggests that it arose de novo). Sanger sequencing of *HCFC1* in 100 normal controls (50 individuals of European descent and 50 African Americans; Human Variation Panel, Coriell Institute) identified a single African American female carrier (NA17132) of the c.344C>T mutation. We concurrently analyzed exome data from 7,597 controls (6,503 from the National Heart, Lung, and Blood Institute [NHLBI] Exome Sequencing Project Exome Variant Server and 1,094 from 1000 Genomes) and did not discover any of the variants detected in our cohort.

HCFC1 contains a kelch domain with five kelch motifs, fibronectin-like domains, and an HCF-proteolysis domain with six HCF repeats[24] (Figure 2). The five mutations identified in the *cblX* individuals affect Gln68, Ala73, and Ala115 within adjacent kelch motifs (Figure 2) and are predicted to be deleterious (Table S4). Furthermore, the predicted three-dimensional structure of wild-type HCFC1 (Figure S2) places Gln68, Ala73, and Ala115 in close proximity to each other. Interestingly, the introduction of the Ala73 and Ala115 substitutions leads to a conformational change in the HCFC1 structure predicted by Modeller[17] (Figures S2B, S2D, and S2E), suggesting that the observed missense mutations could alter the protein structure of HCFC1.

Binding Sites of the THAP11-HCFC1 Complex

ChIP-seq studies in mouse embryonic stem cells[10] have previously shown that the THAP11-HCFC1 complex binds consensus-sequence motifs in genes encoding enzymes of cobalamin metabolism, e.g., *MMACHC*, *MTR*, and *ABCD4* (MIM 603214), as well as *SUCLG1* (MIM 611224). *SUCLG1* encodes a Krebs-cycle enzyme, and mutations in this gene cause a severe form of isolated methylmalonic aciduria.[25] We identified orthologous, conserved binding motifs for the THAP11-HCFC1 complex in human *MMACHC*, *MTR*, *ABCD4*, and *SUCLG1* (Figure S3), suggesting that the expression of human *MMACHC* and possibly other cobalamin metabolism enzymes is regulated by the THAP11-HCFC1 complex.

Analysis of RNA and Protein Expression

To further characterize the functional consequences of the observed *HCFC1* mutations in vitro, we examined the expression of various enzymatic targets—*MMACHC*, *MMADHC*, *MTR*, *SUCLG1*, and *ABCD4*—in skin fibroblasts from two *cblX* individuals (subjects 1 and 11) by using qPCR and immunoblots (Figure 3). Whereas expression levels of *MMADHC*, *MTR*, *SUCLG1* and *ABCD4* in subjects 1 and 11 were similar to those in two control fibroblast lines, *MMACHC* mRNA and MMACHC expression levels (Figures 3A and 3B) were severely reduced in the affected individuals. The expression levels of *HCFC1* mRNA (Figure 3A) and HCFC1 (Figure 3C and Figure S4) were similar in the fibroblasts of control and affected individuals, showing that the missense mutations in *HCFC1* do not affect its expression but apparently inhibit its function in the transcriptional activation of *MMACHC*.

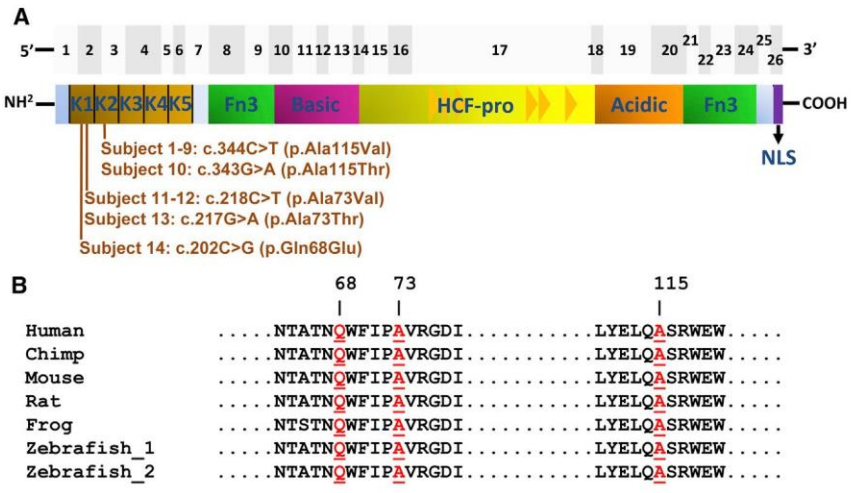

Figure 2. Pathogenic Variants of *HCFC1* in *cblX*

(A) The top panel shows the 26 exons of *HCFC1* as gray boxes. The bottom panel shows the predicted HCFC1 domains, including the kelch domain (kelch motifs K1–K5), Fn3 (fibronectin type 3), the basic domain, HCF-proteolysis repeats (HCF-pro; represented as triangles), the acidic domain, and NLS (nuclear localization signal) domains (adapted from Wilson et al.[24]). The *HCFC1* mutations are clustered within exon 2 and 3 of the cDNA, corresponding to the first (K1) and second (K2) kelch motifs, respectively, in HCFC1. (B) Comparative analysis of HCFC1 from multiple species demonstrated that Gln68, Ala73, and Ala115 (highlighted in red) are evolutionarily conserved throughout vertebrates.

Functional Analysis of *HCFC1* by siRNA Knockdown

In order to further demonstrate the regulatory role of HCFC1 in *MMACHC* transcription, we performed siRNA knockdown of *HCFC1* in HEK293 cells. The reduction in *HCFC1* expression resulted in the coordinate downregulation of *MMACHC* in siRNA-treated cells, but not in untreated control cells (Figure 4). This suggests that *HCFC1* regulates the expression of *MMACHC*, explaining the lower levels of MMACHC expression in the fibroblasts of affected individuals (Figure 3). This is further supported by the observation that fibroblasts from *cblX* individuals produced excessive methylmalonate as a result of reduced MMACHC expression, which normalized after cobalamin supplementation (Figure S5).

Discussion

WES of the index case and his unaffected parents allowed us to discover the mutation underlying his complex clinical phenotype, which included an apparent *cblC* deficiency combined with severe neurological phenotypes. *HCFC1*, unlike other genes associated with cobalamin disorders, does not encode an enzyme in the cobalamin pathway. It is in fact a predicted transcriptional coregulator of enzymes involved in cobalamin metabolism, including that encoded by *MMACHC*, mutations in which are responsible for *cblC* deficiency.[2] The location of *HCFC1* on human chromosome X, combined with the existence of several males with apparent *cblC* deficiency and no *MMACHC* mutations,[2,9] led us to predict that this disorder might represent an X-linked disorder affecting males. Indeed, we identified *HCFC1* mutations in 13/17 additional male subjects with biochemical and cellular findings that were similar to the index case, leading us to designate this cobalamin disorder *cblX*.

All of the mutations discovered in our cohort affect conserved amino acids in two of the five kelch motifs within the HCFC1 kelch domain. Kelch domains are typically repetitive antiparallel β sheets and are known to mediate protein-protein interactions, catalytic activity, and transportation.[26,27] The HCFC1 kelch domain recognizes a conserved HCFC1 binding motif,[28,29] which is important for protein-protein interaction with its transcriptional-regulation partners.[30–33] HCFC1 is known to interact with diverse proteins to regulate a variety of processes, including the cell cycle, proliferation, and transcription,[30,34,35] and exerts control on targets through THAPs, specifically THAP11.[10,11] Thus, mutations that interfere with the interaction between HCFC1 and its transcriptional-regulation partners are likely to affect the expression of a wide range of downstream effectors. It is noteworthy that HCFC1 interacts with many other transcription factors, such as E2Fs, MLLs, CREB, BAP1, and SETD1A, through the kelch domain and subsequently has an extensive effect on transcription activities.[29,30,32,34] In a recent study, the HCFC1 complex was shown to be bound to ~5,400 (24%) promoters of the human transcriptome, including the *THAP11* promoter.[36]

The most obvious question that we needed to address next was how mutations in *HCFC1* might affect cobalamin metabolism and the apparent *cblC* deficiency observed in the *cblX* individuals. We first demonstrated that predicted binding motifs for the THAP11-HCFC1 complex exist in the regions flanking human *MMACHC*, as well as other genes encoding components of the cobalamin pathway. We then showed that expression levels of both *MMACHC* mRNA and MMACHC were severely reduced in skin fibroblasts from *cblX* individuals. Furthermore, siRNA knockdown of *HCFC1* expression in HEK293 cells led to coordinate downregulation of *MMACHC* transcription. Thus, our data strongly support a model in which the THAP11-HCFC1 complex controls the expression of *MMACHC*. Mutations that affect binding of HCFC1 and THAP11 could then lead to the cellular and biochemical phenotype of *cblC* deficiency as a result of a reduction in the expression of MMACHC.

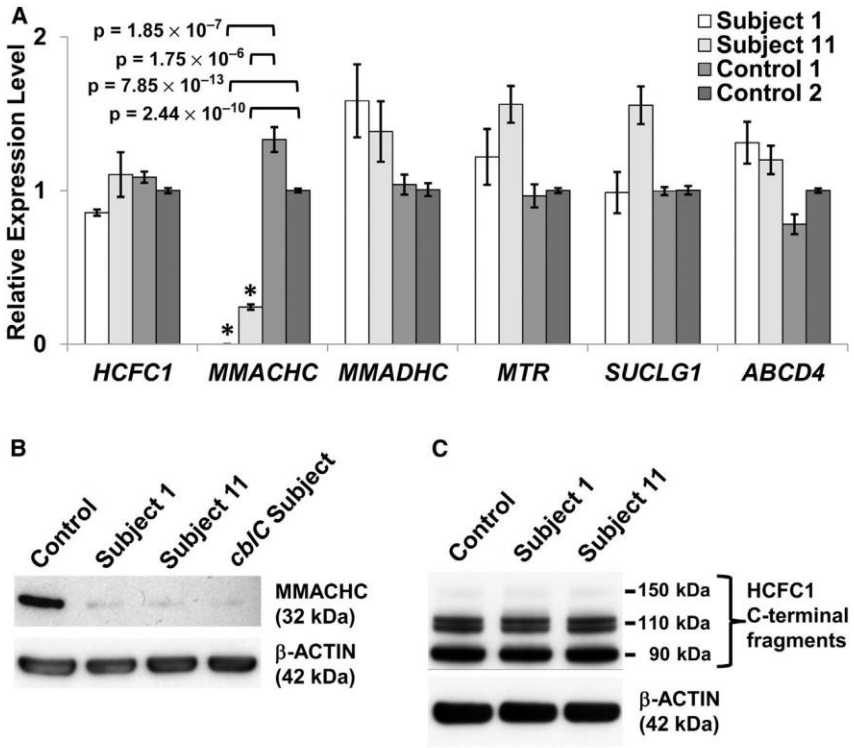

Figure 3. Expression Analysis
(A) qPCR analysis of mRNA expression. *MMACHC* expression was either completely lost or reduced by ~76% in fibroblasts derived from subjects 1 and 11, respectively (asterisks indicate statistical significance). Error bars represent the SEM of relative expression levels. The two control samples used were from healthy individuals with no known biochemical or neurological phenotypes.
(B) Immunoblot analysis of MMACHC. Fibroblast lysates from a reference (control) sample (human dermal fibroblast C-013-5C, Life Technologies), two individuals with *cblX*, and one with *cblC* were analyzed. The *cblX* (lanes 2 and 3) and *cblC* (lane 4) lines show only trace amounts of MMACHC (top panel).
(C) Immunoblot analysis of HCFC1. Fibroblast lysates from the same control used in (B) and two *cblX* cell lines were analyzed. HCFC1 levels in all the samples remained unchanged. β-actin was used as a loading control in all immunoblots (in the bottom of B and C).

The discovery of mutations in *HCFC1* as the cause of the *cblX* disorder highlights perturbation of transcription as the cause of a classical inborn error of metabolism. Although *cblC* and *cblX* share some clinical features, *cblX* individuals were not documented to have the specific bulls-eye maculopathy, which is a frequent finding in *cblC*.[5] Furthermore, the neurological features were more severe (e.g., brain malformation, infantile spasms, movement disorders) in *cblX* individuals, suggesting that *MMACHC* deficiency alone does not explain all of the clinical manifestations. The pathophysiology underlying the complex phenotype remains to be fully elucidated but most likely involves dysregulation of other HCFC1 targets. The recent discovery of variants in the regulatory regions of *HCFC1* as a possible cause of nonsyndromic intellectual disability has implicated *HCFC1* in brain development and function[23] and, together with the data presented here, suggests that the metabolic manifestations of *MMACHC* deficiency, along with mutation analysis of *HCFC1*, should be assessed in individuals with idiopathic X-linked intellectual disability.

We conclude that missense mutations affecting the kelch domain of HCFC1 lead to combined methylmalonic acidemia and hyperhomocysteinemia, establishing a functional relationship between *HCFC1* and cobalamin metabolism. This distinct disorder establishes locus heterogeneity for *cblC* deficiency, most likely mediated by the transcriptional regulation of genes involved in cobalamin metabolism by HCFC1, and demonstrates that an inborn error of metabolism can be caused by transcriptional dysregulation.

Supplemental Data

Supplemental Data include five figures and five tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We are grateful to Renata Gallagher, Janet Thomas, Laura Pickler, and Cynthia Freehauf for their input and advice on the clinical aspects of this project and for their contribution to the care of

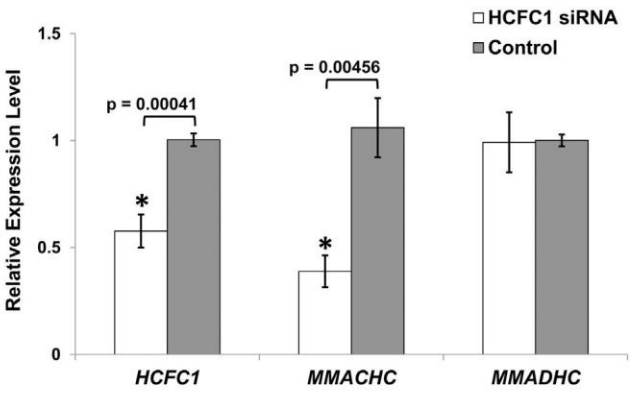

Figure 4. siRNA Knockdown of HCFC1
HEK293 cells were transfected with *HCFC1*-specific siRNA or scrambled siRNA (control), and the relative expression of *HCFC1* and *MMACHC* was assayed by qPCR using *ACTB* (β-actin) as an endogenous control. Compared to control cells (average expression levels were fixed at 1), siRNA-treated cells had significantly reduced *HCFC1* (0.275) and *MMACHC* (0.424) expression levels, but the *MMADHC* expression level remain unchanged. Error bars represent the SEM of relative expression levels.

affected individuals, to Leah Ladores for biochemical and complementation analysis, and to the referring physicians. This work was supported in part by a National Institutes of Health grant (GM081519) to T.H.S. and funds and services from the Colorado Intellectual and Developmental Disabilities Research Center. A.Q. was supported by a grant for postdoctoral research training (T32MH015442), and U.S. was supported by the 2012 Pediatrics Student Research Program at Children's Hospital Colorado. J.L.S., N.P.A., I.M., and C.P.V. were supported by the Intramural Research Program of the National Human Genome Research Institute. D.S.R. was supported by an operating grant from the Canadian Institutes of Health Research (CIHR-M08-15078). M.R.B and B.F. were supported by a grant from the Swiss National Science Foundation (31003A_138521).

Received: May 8, 2013
Revised: July 9, 2013
Accepted: July 26, 2013
Published: September 5, 2013

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/
Burrows-Wheeler Aligner, http://bio-bwa.sourceforge.net/
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
ENZYME, http://enzyme.expasy.org/
Galaxy, http://main.g2.bx.psu.edu/
Integrative Genomics Viewer, http://www.broadinstitute.org/igv/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
SAMtools, http://samtools.sourceforge.net/
SeattleSeq, http://snp.gs.washington.edu/SeattleSeqAnnotation134/
UCSC Genome Browser, http://genome.ucsc.edu/

References

1. Watkins, D., and Rosenblatt, D.S. (2011). Inborn errors of cobalamin absorption and metabolism. Am. J. Med. Genet. C. Semin. Med. Genet. *157*, 33–44.
2. Lerner-Ellis, J.P., Anastasio, N., Liu, J., Coelho, D., Suormala, T., Stucki, M., Loewy, A.D., Gurd, S., Grundberg, E., Morel, C.F., et al. (2009). Spectrum of mutations in *MMACHC*, allelic expression, and evidence for genotype-phenotype correlations. Hum. Mutat. *30*, 1072–1081.
3. Martinelli, D., Deodato, F., and Dionisi-Vici, C. (2011). Cobalamin C defect: natural history, pathophysiology, and treatment. J. Inherit. Metab. Dis. *34*, 127–135.
4. Carrillo-Carrasco, N., Chandler, R.J., and Venditti, C.P. (2012). Combined methylmalonic acidemia and homocystinuria, *cblC* type. I. Clinical presentations, diagnosis and management. J. Inherit. Metab. Dis. *35*, 91–102.
5. Carrillo-Carrasco, N., and Venditti, C.P. (2012). Combined methylmalonic acidemia and homocystinuria, *cblC* type. II. Complications, pathophysiology, and outcomes. J. Inherit. Metab. Dis. *35*, 103–114.
6. Coelho, D., Suormala, T., Stucki, M., Lerner-Ellis, J.P., Rosenblatt, D.S., Newbold, R.F., Baumgartner, M.R., and Fowler, B. (2008). Gene identification for the *cblD* defect of vitamin B_{12} metabolism. N. Engl. J. Med. *358*, 1454–1464.
7. Rutsch, F., Gailus, S., Miousse, I.R., Suormala, T., Sagné, C., Toliat, M.R., Nürnberg, G., Wittkampf, T., Buers, I., Sharifi, A., et al. (2009). Identification of a putative lysosomal cobalamin exporter altered in the *cblF* defect of vitamin B_{12} metabolism. Nat. Genet. *41*, 234–239.
8. Coelho, D., Kim, J.C., Miousse, I.R., Fung, S., du Moulin, M., Buers, I., Suormala, T., Burda, P., Frapolli, M., Stucki, M., et al. (2012). Mutations in *ABCD4* cause a new inborn error of vitamin B_{12} metabolism. Nat. Genet. *44*, 1152–1155.
9. Lerner-Ellis, J.P., Tirone, J.C., Pawelek, P.D., Doré, C., Atkinson, J.L., Watkins, D., Morel, C.F., Fujiwara, T.M., Moras, E., Hosack, A.R., et al. (2006). Identification of the gene responsible for methylmalonic aciduria and homocystinuria, *cblC* type. Nat. Genet. *38*, 93–100.
10. Dejosez, M., Levine, S.S., Frampton, G.M., Whyte, W.A., Stratton, S.A., Barton, M.C., Gunaratne, P.H., Young, R.A., and Zwaka, T.P. (2010). Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes Dev. *24*, 1479–1484.
11. Parker, J.B., Palchaudhuri, S., Yin, H., Wei, J., and Chakravarti, D. (2012). A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function. Mol. Cell. Biol. *32*, 1654–1670.
12. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
13. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
14. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
15. Goecks, J., Nekrutenko, A., and Taylor, J.; Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.
16. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. *38*(Web Server issue), W695–W699.
17. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics *Chapter 5*, 6.
18. Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T.N. (2010). CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. Nucleic Acids Res. *38*(Web Server issue), W576–W581.
19. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.
20. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

21. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277.
22. Sloan, J.L., Johnston, J.J., Manoli, I., Chandler, R.J., Krause, C., Carrillo-Carrasco, N., Chandrasekaran, S.D., Sysol, J.R., O'Brien, K., Hauser, N.S., et al.; NIH Intramural Sequencing Center Group. (2011). Exome sequencing identifies *ACSF3* as a cause of combined malonic and methylmalonic aciduria. Nat. Genet. *43*, 883–886.
23. Huang, L., Jolly, L.A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubridge, C., Wieczorek, D., Tzschach, A., Cohen, M., et al. (2012). A noncoding, regulatory mutation implicates *HCFC1* in nonsyndromic intellectual disability. Am. J. Hum. Genet. *91*, 694–702.
24. Wilson, A.C., Boutros, M., Johnson, K.M., and Herr, W. (2000). HCF-1 amino- and carboxy-terminal subunit association through two separate sets of interaction modules: involvement of fibronectin type 3 repeats. Mol. Cell. Biol. *20*, 6721–6730.
25. Chandler, R.J., and Venditti, C.P. (2005). Genetic and genomic systems to study methylmalonic acidemia. Mol. Genet. Metab. *86*, 34–43.
26. Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., and Knowles, P.F. (1991). Novel thioether bond revealed by a 1.7 A crystal structure of galactose oxidase. Nature *350*, 87–90.
27. Adams, J., Kelso, R., and Cooley, L. (2000). The kelch repeat superfamily of proteins: propellers of cell function. Trends Cell Biol. *10*, 17–24.
28. Lu, R., Yang, P., O'Hare, P., and Misra, V. (1997). Luman, a new member of the CREB/ATF family, binds to herpes simplex virus VP16-associated host cellular factor. Mol. Cell. Biol. *17*, 5117–5126.
29. Freiman, R.N., and Herr, W. (1997). Viral mimicry: common mode of association with HCF by VP16 and the cellular protein LZIP. Genes Dev. *11*, 3122–3127.
30. Tyagi, S., Chabes, A.L., Wysocka, J., and Herr, W. (2007). E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. Mol. Cell *27*, 107–119.
31. Knez, J., Piluso, D., Bilan, P., and Capone, J.P. (2006). Host cell factor-1 and E2F4 interact via multiple determinants in each protein. Mol. Cell. Biochem. *288*, 79–90.
32. Machida, Y.J., Machida, Y., Vashisht, A.A., Wohlschlegel, J.A., and Dutta, A. (2009). The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. J. Biol. Chem. *284*, 34179–34188.
33. Mazars, R., Gonzalez-de-Peredo, A., Cayrol, C., Lavigne, A.-C., Vogel, J.L., Ortega, N., Lacroix, C., Gautier, V., Huet, G., Ray, A., et al. (2010). The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias. J. Biol. Chem. *285*, 13364–13371.
34. Wysocka, J., Myers, M.P., Laherty, C.D., Eisenman, R.N., and Herr, W. (2003). Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes Dev. *17*, 896–911.
35. Yu, H., Mashtalir, N., Daou, S., Hammond-Martel, I., Ross, J., Sui, G., Hart, G.W., Rauscher, F.J., 3rd, Drobetsky, E., Milot, E., et al. (2010). The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. Mol. Cell. Biol. *30*, 5071–5085.
36. Michaud, J., Praz, V., James Faresse, N., Jnbaptiste, C.K., Tyagi, S., Schütz, F., and Herr, W. (2013). HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. Genome Res. *23*, 907–916.