RESEARCH

3D interactions with the growth hormone locus in cellular signalling and cancer-related pathways

Lekha Jain, Tayaza Fadason, William Schierding, Mark H Vickers, Justin M O’Sullivan and Jo K Perry

Liggins Institute, University of Auckland, Auckland, New Zealand

Correspondence should be addressed to J K Perry or J M O’Sullivan: j.perry@auckland.ac.nz or justin.osullivan@auckland.ac.nz

Abstract

Growth hormone (GH) is a peptide hormone predominantly produced by the anterior pituitary and is essential for normal growth and metabolism. The GH locus contains five evolutionarily related genes under the control of an upstream locus control region that coordinates tissue-specific expression of these genes. Compromised GH signalling and genetic variation in these genes has been implicated in various disorders including cancer. We hypothesised that regulatory regions within the GH locus coordinate expression of a gene network that extends the impact of the GH locus control region. We used the CoDeS3D algorithm to analyse 529 common single nucleotide polymorphisms (SNPs) across the GH locus. This algorithm identifies colocalised Hi-C and eQTL associations to determine which SNPs are associated with a change in gene expression at loci that physically interact within the nucleus. One hundred and eighty-one common SNPs were identified that interacted with 292 eGenes across 48 different tissues. One hundred and forty-five eGenes were regulated in trans. eGenes were found to be enriched in GH/GHR-related cellular signalling pathways including MAPK, PI3K-AKT-mTOR, ERBB and insulin signalling, suggesting that these pathways may be co-regulated with GH signalling. Enrichment was also observed in the Wnt and Hippo signalling pathways and in pathways associated with hepatocellular, colorectal, breast and non-small cell lung carcinoma. Thirty-three eQTL SNPs identified in our study were found to be of regulatory importance in a genome-wide Survey of Regulatory Elements reporter screen. Our data suggest that the GH locus functions as a complex regulatory region that coordinates expression of numerous genes in cis and trans, many of which may be involved in modulating GH function in normal and disease states.

Introduction

Growth hormone (GH) is a peptide hormone released in a pulsatile fashion from the anterior pituitary. GH is critical for mediating normal postnatal longitudinal growth in childhood and puberty and in regulating metabolism (Bonert & Melmed 2017). Secretion of GH is positively modulated by GH releasing hormone (GHRH), ghrelin, and negatively by somatostatin. Insulin-like growth factor (IGF-1) is a key mediator of GH actions. Compromised GH/IGF-1 signalling is associated with several well-characterised growth disorders and is linked to an altered

The GH gene cluster is located on the long arm of chromosome 17 (17q23) and is composed of five homologous genes, GH1 (also known as GH-N), GH2 (also known as GH-V) and chorionic somatomammmotropin genes (CSH1, CSH2 and CSHL1) (Liao et al. 2018). Tissue-specific expression of the GH locus genes is regulated by a locus control region which overlaps the CD79B and SCN4A genes that are located upstream of the GH1 gene. GH1 is expressed primarily in the pituitary and other extra-pituitary tissues, whereas the rest of the locus genes are expressed predominantly in the syncytiotrophoblast layer of the placenta (Su et al. 1997). Studies have demonstrated that coordinated regulation of GH locus genes is mediated by complex chromatin looping and epigenetic mechanisms (Kimura et al. 2007, Ganguly et al. 2015, Ho et al. 2008, Tsai et al. 2016).

GH effects are mediated through activation of downstream signalling cascades following binding to the GH receptor (GHR) and through stimulation of secretion of secondary peptide mediator molecules, in particular, IGF-1 (Waters 2016, Bonert & Melmed 2017, Dehkoda et al. 2018). Key signalling pathways pertaining to GH-GHR signal transduction include JAK2 signalling via STATs (1, 3 and 5), the MAPK pathway, JNK pathway, mTOR (mammalian Target of Rapamycin) and PI3K pathway (Carter-Su et al. 2016, Lu et al. 2019). These signal transduction pathways mediate GH effects by altering gene transcription profiles, through direct stimulation of transcription and by modifying chromatin (Rotwein & Chia 2010).

Enhancers and promoters are physically brought together to facilitate the regulation of gene expression by complex 3D mechanisms that are dependent on a multitude of factors (Sanyal et al. 2012, Schierding & O’Sullivan 2015). This physical contact/interaction between enhancers and promoters can be captured by proximity ligation techniques such as Hi-C (Lieberman-Aiden et al. 2009, Eijssouts et al. 2019, Kong & Zhang 2019). A genomic variant associated with allele-specific changes in the expression of a gene is known as an expression quantitative trait locus (eQTL). Notably, the regulation of gene promoters can be mediated through both proximal and distal regulatory elements (cis and trans interactions, respectively), with the latter including interactions between different chromosomes (Gibcus & Dekker 2013, Schierding et al. 2016). These interactions can associate with either higher (enhancer) or lower (insulator/silencer) expression.

As described above, coordinated regulation of GH locus genes is known to be mediated by complex chromatin looping (Kimura et al. 2007, Tsai et al. 2016). In light of recent studies which show regulation of gene networks by alteration of chromosomal interactions in the nucleus (Lanctôt et al. 2007), it is possible that some GH functions may be mediated by spatial interactions between regions of the GH gene locus and distal loci. Changes in chromatin organisation, structure and interactions play a crucial role in the regulation of gene expression (Dekker et al. 2013, Schierding & O’Sullivan 2015, Schierding et al. 2016, Fadason et al. 2017). Therefore, the study of polymorphisms related to genes in this axis could potentially lead to elucidation of novel regulatory networks involving genes associated with GH/IGF-1 axis function.

We hypothesised that regulatory regions within the GH locus coordinate expression of a gene network that extends the impact of the GH locus control region. We integrated 3D genome organisation and tissue specific gene expression data to identify functional cis and trans spatial eQTLs that involved the GH locus. We identified regions within the GH locus that regulate multiple genes involved in key cellular signalling and cancer-related pathways, many of which are related to GH-related signalling pathways.

Materials and methods
Mapping of SNPs across the GH locus
We collated a list of common single nucleotide polymorphisms (SNPs) (dbSNP147; Minor allele frequency ≥1%) located across the GH gene locus including its control region (Chr17:62080000-61920000; GRCh37/hg19) (Supplementary data 1-a, see section on supplementary materials given at the end of this article). SNP density across the GH locus was calculated using a sliding window (500 bp window, 100 bp step size) in RStudio (Version 1.1.414).

Known enhancer sites across the GH locus were obtained from GeneHancer (Fishilevich et al. 2017), which collates information from the ENCODE project (Dunham et al. 2012), Ensembl regulatory build (Zerbino et al. 2018) and FANTOM5atlas of active enhancers.
Topologically associating domains (TADs) within the seven cell lines (i.e. GM12878, HMEC, KBM7, HUVEC, IMR90, K562 and NHEK) were determined using the 3D Genome Browser at 1-kb resolution and Hi-C data from Rao et al. (2014).

Identification of eQTLs within the GH locus and their genome-wide targets

GH locus SNPs were analysed using the CoDeS3D (Contextualize Developmental SNPs using 3D Information) algorithm (GitHub, https://github.com/Genome3d/codes3d-v1) (Fadason et al. 2017). CoDeS3D integrates genome spatial connectivity data (i.e. maps of loci that physically interact, captured by Hi-C (Rao et al. 2014)) and links it to eQTL data obtained from the Genotype-Tissue Expression (GTEx) database (v7) (Ardlie et al. 2015).

These results were corrected for false discovery using the Benjamini–Hochberg correction procedure (Fadason et al. 2017). The CoDeS3D database was loaded with Hi-C data for GM12878, IMR90, HMEC, NHEK, K562, HUVEC and KBM7 human cell lines (Rao et al. 2014). Cis-eQTLs were defined as involving SNPs that were located <1 Mb from the affected gene (or eGene), whereas trans-eQTLs were defined as involving SNPs located >1 Mb from the affected eGene on the same chromosome or on a different chromosome (Fig. 1A) (Fadason et al. 2017).

Pathway enrichment analyses

To identify enrichment of eGenes within biological pathways, we analysed the eGenes set (identified by CoDeS3D) using gProfileR package in R (https://biit.cs.ut.ee/gprofiler/). The reference gene sets were from KEGG.
3D interactions with the GH locus

Table 1  Overview of connections in cis and trans identified by CoDes3D.

<table>
<thead>
<tr>
<th>SNPs analysed</th>
<th>SNPs with eQTLs</th>
<th>eGenes</th>
<th>Cis eGenes</th>
<th>Trans eGenes</th>
<th>Cis connections</th>
<th>Trans connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>529</td>
<td>181</td>
<td>292</td>
<td>32</td>
<td>260</td>
<td>2141</td>
<td>708</td>
</tr>
</tbody>
</table>

(KEGG FTP Release 2019-09-30), Reactome (annotations: ensemble classes: 2019-10-2) and WikiPathways (20190910). An adjusted P value <0.05 was significant following Benjamini–Hochberg correction for false discovery rate (FDR) (Raudvere et al. 2019).

Data visualisation

Figures were drawn in RStudio (Version 1.1.414) using the following packages/libraries:

CircOS (Gu et al. 2014), ggplot2 (Gómez-Rubio 2017) and magrittr (Bache & Wickham 2016).

Results

Common polymorphisms across the GH gene locus are associated with expression of multiple downstream genes

We used CoDeS3D to analyse 529 common SNPs (dbSNP 147) across the GH locus (Chr17:62080000-61920000; GRCh37/hg19) and identify SNP-eGene pairs in which the SNP was associated with the eGene expression level (Fig. 1B and Supplementary data 1-a) (Fadason et al. 2017). This analysis identified 181 SNPs that interacted with 292 genes in 48 different tissues (Table 1 and Supplementary data 1-b). There were 2141 SNP-eGene associations (FDR <0.05, Benjamini–Hochberg correction) in cis (<1 Mb distance between the SNP and eGene) and 708 associations in trans (where the SNP and eGene are >1 Mb apart or on different chromosomes).

We compared the distribution of the eQTL SNP-eGene interaction frequency and query - SNP density across the selected region (Fig. 1B). Notably, the number of SNPs per 500 bp sliding window did not correlate ($R^2$=0.1) with the number of functional eQTL-eGene interactions (Supplementary Fig. 1). Thus, the identification of regions with functional eQTL-eGene interactions was not an artefact of regions of higher SNP density across the GH locus. However, the number of cis-eQTLs in each tissue correlated positively ($R^2$=0.64) with the number of samples present in GTEx for the respective tissue (Supplementary Fig. 2). Notably, there are some tissues that are outliers in the correlations, including the pituitary, cerebellum, oesophagus, pancreas, breast and adipose tissues. Interestingly, most of these tissues have key functions that relate to GH action (Bartke 2011, Clasen et al. 2014, Duan et al. 2015, Tarnawski et al. 2015, Bonert & Melmed 2017, Pekic et al. 2017).

Within the nucleus, chromosomes are arranged through a hierarchy of structures that include topologically associating domains (TADs). TADs are defined as regions where spatial contacts are enriched (Tang et al. 2015). To determine whether the observed SNP-eGene connections crossed TAD boundaries, a Hi-C heat map was generated at 1-kb resolution using Hi-C data captured in the GM12878, HMEC, KBM7, HUVEC, IMR90, K562 and NHEK cell lines (Rao et al. 2014). The Hi-C analysis clearly showed that there were numerous SNP-eGene connections observed both within the individual TAD containing the GH locus region and across the TAD boundaries (Fig. 2A and Supplementary Fig. 3). This is consistent with SNP-eGene connections not being limited to occurring within TADs (Ciabrelli & Cavalli 2015, Ulianov et al. 2016, Chen et al. 2018). Connections were also observed with eGenes present on different chromosomes (Fig. 2B), consistent with the concept that physical interactions between two genomic regions are not restricted by proximity in linear distance.

As the SNPs which we used for our study were part of the GH region (which consists of SCN4A, CD79B, GH1, CSHL1, CSH1, GH2, CSH2 and TCAM1P genes), we analysed whether there were any connections going back into these genes. SNPs (n=73) located across the entire GH region had connections with CD79B in 16 tissues, one SNP with CSHL1, three SNPs with GH2 in pituitary tissue, 114 SNPs with CSH2 in three tissues and 173 SNPs with TCAM1P gene in 24 tissues. As the locus control region is critical for regulation of these genes, we focused on interactions in this region. Twenty-seven SNPs from across the GH locus control region connected to CD79B in 13 tissues, one SNP to CSHL1 and one SNP to GH2 in pituitary tissue, 19 SNPs to CSH2 in three tissues (testis, cerebellum and cerebellar hemisphere) and 33 SNPs to TCAM1P gene in 18 tissues (Supplementary data 1-b). This is consistent with Tsai et al. (2016) who used chromatin conformation capture (3C) data from human pituitary and placental tissues to demonstrate that the GH locus control region regulates these genes (Tsai et al. 2016). Notably, we identified only a few eQTLs with CSHL1 and GH2 and no eQTLs with the GH1 gene, which may reflect...
Figure 2

(A) Distribution of topologically associating domain (TAD) structures with (A) cis-eGenes (i.e. eGenes located <1 Mb from the SNP). The Hi-C heat map was generated with the 3D Genome Browser using data from Rao et al. (2014) for all the seven cell lines (GM12878, HMEC, KBM7, HUVEC, IMR90, K562 and NHEK). Tracks show TAD structures, the region containing the SNPs and identified eGenes. Blue and yellow bars represent different TAD regions. (B) A circus plot illustrating all the connections from GH locus SNPs to genes present on different chromosomes which includes chromosome 17. These eQTLs were not just limited to chromosome 17 but extended to multiple chromosomes. The circus plot has been split into two sections with the green curve representing the GH locus region across which the SNPs were analysed.
the age, sex (65.8% male, 68.5% above 50 yrs of age) and tissue distribution (i.e. no placental samples) that was used to characterise expression profiles within the GTEx database (Ardlie et al. 2015).

CD79B displays a very complex eQTL-pattern. Eight CD79B SNPs (rs1051684, rs12603821, rs1051688, rs12451467, rs2070776, rs2005132, rs2320125 and rs8077653) connect to 66 eGenes in 48 different tissues.

Figure 3
Pattern of cis-eQTL connections in different tissues. Identified contacts with genes present in the local environment of the GH gene locus appear to cluster together in four different pattern types as shown in the figure. Representative figures for patterns A, B, C and D are the connections observed in omental visceral adipose, cortex, skeletal muscle and transverse colon, respectively. Rectangles in black represent the genes of GH locus across which SNPs were analysed and those in grey are the associated cis-eGenes. Links in red represent upregulation of the eGene associated with the SNP, whereas the links in black represent downregulation.
Out of these genes, 47 eGene contacts were in trans (>1 Mb away, distal) and 19 were in cis (<1 Mb away, nearby). Out of 47 trans eGenes, 24 genes were distal but present on chromosome 17 (up to 59 Mb downstream and 18 Mb upstream from CD79B). These eight SNPs were also associated with increased expression of the CD79B gene in 11 tissues and a decrease in CD79B expression in atrial appendage of the heart. Seventy-three output SNPs from across the analysed region were found to be associated with altered CD79B expression in 16 tissues. Some of these SNPs were linked to an increase in expression of CD79B, whereas some were linked to the downregulation of this gene in a tissue-specific manner. This differential expression across tissues could imply disruption of binding sites of tissue-specific transcription factors by these or linked SNPs. For example, rs3815358 and rs12452767 occur in a CTCF binding site. CTCF is an important transcriptional regulator protein (Splinter et al. 2006).

Disruption of these CTCF binding sites can block or cause inefficient binding of transcription factors or may impact on other regulatory processes (Ohiolson et al. 2001).

To assess the cis connections with the GH locus SNPs, we investigated eQTLs with genes immediately downstream of the GH locus- SMARC2D2, PSMC5, FTSJ3, DDX42, CCDC47, STRADA, LIMD2 and MAP3K3. We observed four strikingly different normalised effect size (NES) patterns in eQTL associations involving these genes across all the tissues (Fig. 3). Eleven tissues were classified as pattern A, twelve as pattern B, four as pattern C and fifteen as pattern D. It was interesting to observe such a marked difference between these patterns; however, the functional significance of these differences remains to be determined. Collectively, our results are consistent with the presence of regulatory sites located within the GH locus.

**Regions in the GH locus interact with multiple genes (cis and trans) in a tissue-specific manner**

Within the GH locus, CSH2 exhibited the maximum number of associations with 114 SNPs in three tissues – testis, brain cerebellum and brain cerebellar hemisphere. In addition, the ICAM2 and FTSJ3 genes associated with 108 SNPs in one tissue (skeletal muscle) and 103 SNPs across all 48 tissues, respectively (Supplementary data 1-c). CSH2 is part of the GH locus and has a high sequence similarity with other GH genes.

While the cis-eGenes mostly had eQTLs with multiple SNPs in several tissues, trans-eGenes exhibited more tissue-specificity (Supplementary data 1-d.1). One hundred and thirty-nine eGenes formed an eQTL with only one SNP, whereas 55 had associations with two SNPs. Conversely, seven genes (SCPEP1, B3GNT1I, NCCOR1, RGS9, VMP1, LINC00511 and CACNG4) had eQTLs with multiple SNPs (Supplementary data 1-d.2). For example, CACNG4...
(Calcium Voltage-Gated Channel Auxiliary Subunit Gamma 4) was found to be associated with 40 SNPs in the atrial appendage.

Interestingly, we observed differences in the proportions of cis and trans eQTLs in different tissues. For example, the pituitary, which had the highest number of cis-eQTLs compared to all the tissues, did not have the highest number of trans-eQTLs. Brain-related tissues had the most eQTL-associations in trans (Supplementary data 1-e), whereas the pituitary had the maximum number of eQTLs with genes in cis (Supplementary data 1-f). This is particularly notable considering the role of the pituitary in GH-related function.

Some SNPs such as rs3815358 are connected with multiple eGenes (27 eGenes) in a variety of tissues (37 tissues), whereas a few SNPs such as rs11869827 are only associated with one or two eGenes (Supplementary data 1-g and h). The maximum number of eGenes which SNPs in our data were associated with was 27. More than 50% of the SNPs exhibited eQTL associations in all 48 tissues, which could imply that these SNPs could potentially have a role in regulation of normal cell functioning (Supplementary data 1-i).

### Genes targeted by GH eQTLs are potentially involved in growth hormone functions

To explore if the set of eGenes (Supplementary data 1-j) were enriched for specific signalling pathways, we analysed the eGene sets using g:Profiler (Raudvere et al. 2019) (Supplementary data 1-k and l). The top 20 most significant (adjusted P value) enriched pathways are summarised in Fig. 4. This analysis demonstrated enrichment for a subset of these genes in numerous GH-related cellular signalling pathways such as the PI3K-AKT, MAPK, mTOR, prolactin, insulin and ERBB signalling pathways (Supplementary data 1-k and l). Another pathway which is potentially involved in regulation of GH signalling is the Wnt pathway, which was also among the enriched pathways (Supplementary data 1-m). There was also a significant representation of these genes in carcinogenic pathways.

Five cancer terms were enriched in the KEGG subset and seven in the WikiPathways subset. There were four cancer types in common in both sets; these were hepatocellular carcinoma, colorectal cancer, breast cancer and non-small cell lung carcinoma. Altered GH signalling has previously been established to be associated with multiple cancer states, including the ones we identified (Chhabra et al. 2011, Clayton et al. 2011, Petry et al. 2017). There was also eGene enrichment in other pathologic conditions as well, like Alzheimer’s and Huntington’s disease. Altogether, these observations are consistent with the hypothesis that genetic variation in the GH locus is associated with modulation of GH function.

### Identified eQTLs co-localise with GWAS signals

To determine if the eQTLs we identified are associated with a disease or a population, we cross-referenced our list of SNPs with the genome-wide association studies (GWAS) catalogue (Buniello et al. 2019) and found that five variants (rs2005172, rs2070776, rs2532111, rs28386778 and rs2854160) exhibited GWAS associations with certain traits or populations (Table 2). rs2005172, rs2070776 and rs2532111 are associated with fat-free mass, height and waist-hip ratio, respectively, in individuals of European ancestry. SNP rs2854160 is associated with height in East Asian and African population, whereas rs28386778 is linked to prudent dietary patterns in 141 individuals. This is important in the context of GH biology, since regulation of height is one of the major functions modulated by this hormone. Excess or deficit in GH can lead to multiple growth disorders such as dwarfism, gigantism and acromegaly. SNP rs2070776, which is linked to height in European population, connects to 26 eGenes, (11 genes in cis and 15 in trans) across 39 different tissues. Six of the trans-eGenes are located on different chromosomes. One of these identified eGenes, NFI (Neurofibromin 1), which is more than 1 Mb distance away from rs2070776, is known to be strongly related to human height (Kehrer-Sawatzki et al. 2017). The height-related SNP rs2854160 in individuals of East Asian and African descent connects to 13 eGenes in 42 tissues in which there are 11 cis-eGenes and two trans-eGenes (on the same chromosome). These two SNPs are located in different regions in the GH locus.

### Regulatory potential of GH eQTLs are supported by functional studies

We next sought experimental evidence for the SNP eQTL-eGene connections from published literature.

van Arensbergen et al. recently developed a functional screening method known as Survey of Regulatory Elements (SuRE) which identifies SNPs that impact on regions of regulatory significance (van Arensbergen et al. 2019). Using this method, a total of 5.9 million SNPs were surveyed in two cell lines – human erythroleukemia cells (K562) and human hepatocellular carcinoma cells (HepG2) – to identify SNPs that alter the activity of putative regulatory elements (van Arensbergen et al. 2019). Thirty-three of the
Figure 5

eGenes identified by CoDeS3D enriched in the mTOR (A) and Wnt (B) signalling pathways. These figures illustrate where these pathway-related eGenes fit in the cellular signalling cascade. Genes/proteins in orange boxes represent eQTL pathway-related genes, whereas those in blue boxes are non-eQTL genes. The eGenes and tissues in which these eQTLs were identified and the SNPs in which they connected to are summarised in Supplementary data 1-r and q.
SNPs identified in our study were demonstrated to be of regulatory importance by van Arensbergen et al. Sixteen SNPs were identified in HepG2 cells and 16 in K562 cells, whereas one SNP, rs12451467, was found to be significant both cell lines (Supplementary data 1-n). Notably, our results also revealed Hi-C interactions between 17 SNPs with eGenes in K562. For example, SNP rs2584608 has regulatory interactions with 12 eGenes (including the chromatin-remodelling factor SMARCD2 and papillary thyroid carcinoma biomarker LIMD2), across 48 different tissues (Supplementary data 1-o).

To further strengthen the functional application of our approach, we analysed known enhancer/promoter regions using data retrieved from GeneHancer. There were 16 enhancer/promoter regions in the region used for our study (Chr17:62080000-61920000; GRCh37/19) which target 13 of the identified eGenes (Supplementary data 1-p). These regulatory regions encompass seven of the output SNPs (rs12452767, rs2286564, rs2286565, rs2457681, rs2665808, rs2854184, rs34684062, rs3815358, rs6171 and rs8080613), out of which two SNPs (rs2286564, rs2286565) were shown to have functional regulatory relevance in the data shown (Supplementary data 1-n). Collectively, these observations provide support for the functional roles of the eQTLs identified in this study.

Discussion

We analysed common SNPs within the GH locus and identified it as a potential regulatory hub for genes located not only on chromosome 17 but also on other chromosomes. Many of the identified genes are a part of GH-related cellular signalling pathways including pathways in cancer. These results suggest that some GH functions could potentially be mediated by interaction of regulatory regions within the GH locus region.

When we examined the specifics of these eQTL-eGene interactions, we observed interesting gene expression patterns in a distinct tissue-specific manner. One hundred and fourteen SNPs from the entire GH region were found to be associated with altered regulation of gene expression of CSH2 gene in the testis and brain tissues (cerebellar hemisphere and cerebellum). It should be noted that samples named brain cerebellum and brain cerebellar hemisphere in GTEx are considered duplicates, as they were the same tissue taken at different times post-mortem (Ardlie et al. 2015). The presence of the eQTL in both indicates that the mRNA was not subject to rapid degradation due to senescence. Most of the CSH2 associated eQTL SNPs were found clustered just downstream of CSH2 gene. Although CSH2 mRNA expression is extremely low in the testis and brain (GTEx), CSH2 protein expression is observed in some germ cell tumours of the testis (Berger et al. 1999). CSH2 has also been linked to foetal growth disorders, pre-eclampsia and choriocarcinoma (Kim 2003, Männik et al. 2010, Liu et al. 2011).

CD79B is an immune-related gene which is important for initialising signal transduction activated by the B-cell receptor complex (Alfarano et al. 1999). Polymorphisms in this gene have been linked to several types of cancer. Mutations in CD79B have been linked to different types of diffuse large B-cell lymphoma (DLBCL) (Frick et al., 2018, Schrader et al. 2018) and to osteosarcoma (Mirabell et al. 2011). Identified SNPs in CD79B and CSH2 have enhancer marks in multiple human tissues and cell-lines. Analysis of known regulatory regions from GeneHancer showed that GH17J063930 is an enhancer for both CD79B and CSH2 and GH17J063877 is an enhancer for CSH2. This super-enhancer region encompasses four SNPs (rs2286564, rs2286565, rs3815358 and rs12452767), and the CSH2 enhancer overlaps with one SNP, rs2457681. Consistent with this, all four SNPs had eQTLs with CD79B and CSH2 in our analysis and rs2457681 connected with CSH2 (Supplementary data 1-b). The study by van Arensbergen et al., which validated SNPs important for regulatory activities, identified two of the SNPs from this enhancer region (rs2286564 and rs2286565) as variants which are of regulatory importance (van Arensbergen et al. 2019). This is consistent with our hypothesis that the entire GH locus potentially serves as a complex regulatory region.

Pathway analysis (g:Profiler) of the identified eGenes (292 genes) identified enrichment of these genes in various cellular signalling pathways, including classical GH-related pathways such as the PI3K-AKT, mTOR, MAP kinase, prolactin and insulin signalling pathways. The role of mTOR signalling in growth and development and mediation of GH function is well-established. mTOR is a protein kinase that regulates cell metabolism, proliferation and survival and is important for mediating GH-related pathways, including those related to insulin and IGF actions (Hayashi & Proud 2007, Bartke 2011, Saxton & Sabatini 2017), so it was interesting to find that eGenes associated with SNPs from the locus were enriched in the mTOR pathway (Fig. 5A; Supplementary data 1-q). Another pathway enriched in this analysis was the Wnt signalling cascade, which mediates embryonic development pathways, cell polarity, migration and division (Fig. 5B; Supplementary data 1-r). The enriched eGenes were part of both canonical and non-canonical Wnt signalling.
as there is significant crosstalk between Wnt and other signalling pathways like mTOR, MAP kinase and P13K, it is possible that there was just an overlap between GH-mediated signalling and the Wnt pathway. However, several studies suggest a potential connection between GH and Wnt signalling (Vouyovitch et al. 2016, Osmundsen et al. 2017). For example, Vouyovitch et al. demonstrated that GH regulates the expression of the secreted protein Wnt4, a Frizzled receptor ligand (Vouyovitch et al. 2016). Therefore, we contend that GH may mediate some of its functions through direct impact on gene expression levels of the Wnt signalling cascade components, and this may be mediated by spatial interaction of these regions with the GH locus. To confirm this, putative regulatory regions can be confirmed with reporter assays, and single-nucleotide editing techniques such as CRISPR/Cas coupled with proximity ligation assays like Hi-C and gene expression studies may substantiate the impact of allele variants on the expression of Hi-C linked genes in vitro or in animal models. However, studies of this nature are complex, as regulatory regions often act in a combinatorial manner and, therefore, modification of isolated SNPs may have minimal impact.

Pathway enrichment analysis identified that eGenes associated with GH locus SNPs were overrepresented in pathological conditions such as cancer, growth disorders and Alzheimer’s disease. There is extensive literature supporting a role for GH in cancer. Altered GH signalling and increased expression of GH and IGF-1 is linked to progression of numerous cancer types (Perry et al. 2013, Simpson et al. 2017, Lu et al. 2019). Autocrine GH increases the size of hepatocellular tumour xenografts and is associated with a worse relapse-free and overall survival in patients with hepatocellular carcinoma (Kong et al. 2016). GH decreases expression of the tumour suppressor gene p53 in the colon and contributes to the development of colorectal carcinoma (Chesnokova et al. 2016). Elevated GH expression is also observed in colorectal cancer and is positively associated with tumour size and lymph node metastasis (Wang et al. 2017). Similarly, there are multiple studies which have established the key role of GH in breast cancer, endometrial and non-small cell lung carcinoma (Pandey et al. 2008, Chhabra et al. 2011, 2018, Perry et al. 2013, Lu et al. 2019). Identified eGenes in our data show significant enrichment in all of these cancer types, which suggests that the eGene set is strongly associated to GH-related cancer types, not purely by chance. This data suggests that genes in and around GH locus and those associated with polymorphisms across the locus have a direct or indirect association with cancer, possibly by alteration of gene regulatory networks.

We also found an overlap between recent studies which identified genes associated with Alzheimer’s disease and the eGenes in our data. There were five genes (ACE, PRKCA, ERN1, GSK3B and MAP2) from one study (Grimm et al. 2019) and two genes from another (KRTAPS-AS and PSMC5) (Kikuchi et al. 2019) which were shared with the identified eGene set. This could possibly indicate a link between GH gene locus mediated/coordinated gene regulation and Alzheimer’s disease. However, significant experimental evidence is needed to substantiate this.

We show that there is a physical contact between SNPs across the GH locus (including the locus control region) and GH2 and CSHL1 genes. In addition, we identify an association of these SNPs with change in expression levels GH2 and CSHL1 in pituitary tissue. This result can be linked back to studies from the Liebhaber and Cooke laboratories (Kimura et al. 2007, Ho et al. 2008, Tsai et al. 2016), which show that genes of the GH locus (GH1, GH2, CSH1, CSH2 and CSHL1) are expressed in the placenta or pituitary in a tissue-specific manner and are coordinated by the locus control region overlapping genes CD79B and SCN4A. This regulation is under epigenetic control and is affected by chromatin looping that brings these regions into close proximity with the target promoters.

GH is crucial for modulation of normal growth and metabolism in the human body, through stimulation of signalling pathways and regulation of multiple growth factors. It is well-established that GH-related downstream pathways, such as mTOR, and other pathways identified here, such as Wnt signalling, are also pivotal in central biological roles affecting growth, development and metabolism (Taciak et al. 2018, Liu & Sabatini 2020). The impact of GH on critical signalling cascades and their associated functions is likely a consequence of the interplay of genetic, molecular and evolutionary factors. Collectively, our observations support the hypothesis that the GH locus (which includes GH1, GH2, and CSH1, CSH2 and CSHL1 and its locus control region spanning genes CD79B and SCN4) functions as an extended regulatory region that coordinates expression of genes located both within and outside of the locus, which are in key GH-linked pathways and contribute to GH function. This is particularly important to consider in the context of evolutionary biology, as it is unlikely that co-regulation of these pathways has evolved by chance. Instead, it is consistent with the premise that genetic regions that are critical in coordinated regulation of aligned biological
processes are linked as part of their maintenance. In the context of the human GH locus, it is notable that most non-primate mammals only have a single GH gene and that the GH cluster arose from gene duplication independently in New World and Old World monkeys and thus varies considerably in structure (González Alvarez et al. 2006, Wallis & Wallis 2006). Comparative studies of the 3D interactions between the GH locus of primate and non-primate mammals with genes in pathways identified in our study may highlight the evolutionary significance of the co-regulation of these pathways.

Supplementary materials
This is linked to the online version of the paper at https://doi.org/10.1530/JME-20-0010.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
J P, J O S, M V and L J were funded by a University of Auckland FRD grant (3714499).

References


testicular DLBCL and gastrointestinal DLBCL. *Leukemia and Lymphoma* 59 1260–1263. (https://doi.org/10.1080/10428194.2017.1
370546)


Ohlsson L, Renkawitz R & Lobanenkov V 2001 CTCF is a uniquely transcriptional regulator linked to epigenetics and disease. *Trends in Genetics* 17 520–527. (https://doi.org/10.1016/S0168-9525(01)02366-6)


principles of chromatin looping. Cell 159 1665–1680. (https://doi.org/10.1016/j.cell.2014.11.021)