Regulation of p110α PI 3-Kinase Gene Expression

Klaartje Kok1,2, Gemma E. Nock1, Elizabeth A. G. Verrall1*, Michael P. Mitchell3, Daan W. Hommes4, Maikel P. Peppelenbosch2, Bart Vanhaesebroeck1*

1 Centre for Cell Signalling, Institute of Cancer, Queen Mary University of London, Charterhouse Square, London, United Kingdom, 2 Department of Cell Biology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands, 3 Bioinformatics and Biostatistics, Cancer Research UK London Research Institute, London, United Kingdom, 4 Department of Gastroenterology and Hepatology; Leiden University Medical Centre, Leiden, The Netherlands.

Abstract

Background: Despite an intense interest in the biological functions of the phosphoinositide 3-kinase (PI3K) signalling enzymes, little is known about the regulation of PI3K gene expression. This also applies to the leukocyte-enriched p110α catalytic subunit of PI3K, an enzyme that has attracted widespread interest because of its role in immunity and allergy.

Principal Findings: We show that p110α expression is mainly regulated at the transcriptional level. In fibroblasts, lymphocytes and myeloid cells, p110α gene transcription appears to be constitutive and not subject to acute stimulation. 5'RACE experiments revealed that p110α mRNA transcripts contain distinct upstream untranslated exons (named exon -1, -2a, -2b, -2c and -2d), which are located up to 81 kb upstream of the translational start codon in exon 1. The levels of all the different p110α transcripts are higher in leukocytes compared to non-leukocytes, with the p110α transcript containing exon -2a most abundantly expressed. We have identified a highly conserved transcription factor (TF) binding cluster in the p110α gene which has enhanced promoter activity in leukocytes compared to non-leukocytes. In human, this TF cluster is located immediately upstream of exon -2a whilst in mouse, it is located within exon -2a.

Conclusion: This study identifies a conserved PIK3CD promoter region that may account for the predominant leukocyte expression of p110α.

Introduction

Phosphoinositide 3-kinases (PI3Ks) generate lipid second messengers that regulate a broad variety of cellular responses such as growth, cell cycle progression, differentiation, vesicular traffic and cell migration [1]. PI3K activity is critical in a wide variety of normal and pathological physiological responses, including immune regulation, metabolic control and cancer. However, despite the importance of this signalling system, very little is known about the regulation of PI3K gene expression under normal and disease conditions.

The PI3K family is divided into 3 classes [2]. Class I PI3Ks are acutely activated upon receptor stimulation and are heterodimers consisting of a p110 catalytic subunit in complex with a regulatory subunit. The class I PI3Ks are further subdivided into class IA and IB, depending on whether the catalytic subunit is in complex with an SH2-domain containing regulatory subunit (collectively called ‘p85’ or with the p101 or p84 regulatory subunits, which lack SH2 domains. Mammals have 3 class IA p110 catalytic subunits, p110α, p110β and p110δ, encoded by 3 distinct genes, PIK3CA, PIK3CB and PIK3CD, respectively. These p110 isoforms interact with p85, of which there are at least five different species, called p85α, p55γ and p50α (encoded by the PIK3R1 gene) and p85β and p55γ (encoded by PIK3R2 and PIK3R3, respectively). p110γ is the only class IB PI3K catalytic subunit and occurs in complex with p101 [3,4] or p84 [5,6], which have no homology to p85. Class I PI3Ks can be activated by tyrosine kinases (p110α, p110δ) or GPCRs (p110β and p110γ) [1,7–9].

Tissue distribution and the regulation of PI3K expression has recently been reviewed [10]. Whereas p110α and p110β appear to have a broad tissue distribution [11–15], p110δ is highly expressed in leukocytes [12,13,16], found at intermediate levels in neurons [17] and present at low levels in most other cell types [13,18]. p110δ is also expressed at moderate levels in some cancer cells of non-leukocyte origin such as melanoma and breast cancer cells, often with large differences in expression levels in cell lines of the same tissue origin [18], for reasons that are unclear at the moment. Like p110δ, p110γ is highly enriched in leukocytes [19–21] but is also found at lower levels in other cell types such as cardiomyocytes [22–24], endothelial cells [25], pancreatic islets [26,27] and smooth muscle cells [28].

Expression of the class IA catalytic isoforms can be altered during physiological and pathological processes, including differentiation (p110α and p110β) [29], regeneration (p110α) [30,31] and hypertension (p110β and p110δ) [32–34]. PI3K expression, especially of p110α, is also very frequently increased in cancer.
Insulin and nuclear receptor ligands can induce expression of the class I regulatory subunits [35–39]. Other documented mechanisms of p85 regulation are through the transcription factors (TF) STAT3 (p55α and p59α) [40], EBNA-2 (p55α) [41] and SREBP (p55γ) [42] and through targeted degradation of p85α and p85β by microRNAs [43,44] (reviewed in [10]).

Three recent studies have identified a transcription regulatory region for the human p110γ gene, PIK3CD. The PIK3CD locus gives rise to two alternative transcripts which each contain a distinct 5′ untranslated exon (exon -1b or -1a) that is spliced onto the first translated (ATG-containing) exon. The genomic position of these 5′ untranslated exons is about 50 kb upstream of the translation start site [45,46]. TF binding sites for p53 [45], FOXO3α [46] and NF-kB [47] have been mapped in close proximity to the most 5′ untranslated exon (called exon -1b). Whereas p53 might inhibit transcription of p110γ, evidence for a positive regulation by NF-κB and FOXO3α has been presented.

A promoter region for murine p110γ has also been identified [19]. Multiple transcriptional start sites exist for p110γ, resulting in transcripts with varying 5′ untranslated regions (5′UTRs), up to 874 bp in length. Analysis of the genomic p110γ DNA up to 1.2 kb upstream from the transcription start site revealed that the putative promoter region contains consensus sites for housekeeping TFs such as AP1 and SP1, as well as several putative binding sites for leukocyte-specific TFs [19]. Functional analysis of this p110γ putative promoter region revealed enhanced promoter activity in the U937 myeloid cell line compared to the HeLa epithelial cell line [19].

In this study, we have investigated the regulation of p110γ gene expression. We have documented that p110γ protein expression largely correlates with the level of p110γ mRNA in numerous cell types, indicating that p110γ expression is predominantly regulated at the level of transcription. We have found multiple mouse and human p110γ transcripts that contain distinct upstream untranslated exons, which we have named exon -1, -2a, -2b, -2c and -2d, located up to 81 kb upstream of the translational start codon in exon 1. Furthermore, we have identified a highly conserved TF-binding cluster that is located within mouse exon -2a and located immediately 5′ upstream of human exon -2a. This TF-binding cluster has enhanced promoter activity in leukocytes compared to non-leukocytes. Out of the 7 different TF binding sites in the TF-binding cluster, 4 are associated with regulation of hematopoiesis and expression of leukocyte-specific genes. These findings are the first to identify a PIK3CD promoter and offer a rationale for the leukocyte-enriched expression of p110γ.

Results

p110δ protein expression is not altered in fibroblasts, B-lymphocytes and myelomonocytic cells upon acute stimulation with various agonists

We first investigated whether p110δ expression can be induced by several acute cellular stimuli. In NIH-3T3 fibroblasts, which contain very low levels of endogenous p110δ compared to leukocytes, p110δ could not be induced by TNF, the proteasome inhibitor PS-341, UV irradiation, osmotic stress or the glucocorticoid dexamethasone (data not shown). p110δ protein levels were also unaffected during different phases of the cell cycle in these cells (data not shown). In B lymphocytes, p110δ expression was not affected by stimulation of the antigen receptor using anti-IgM antibodies. In U937 myelomonocytic cells, p110δ levels were unaltered by treatment with retinoic acid, in contrast to the p110γ protein which was induced effectively (data not shown), the latter in line with previously published data [48,49]. Taken together, p110δ expression appears not to be regulated in an acute manner in response to extracellular stimuli, at least in the cell types and conditions investigated.

Correlation between p110δ mRNA and protein levels in cell lines

We next assessed the levels of p110δ protein and mRNA, using immunoblotting of total cell lysates and real time RT-PCR, respectively, in a panel of murine and human cell lines (Figure 1A,B). p110δ mRNA and protein were found in all cell lines investigated but in widely varying amounts. In line with published data [12,13,16,18], leukocytes expressed high levels of p110δ while non-leukocytes expressed intermediate to low levels. In line with previous data [12], a good correlation was found between p110δ mRNA and protein levels in most cell lines tested, indicating that p110δ protein expression is mainly regulated at the level of transcription.

DNA methylation and histone acetylation are unlikely to be key mechanisms to control PIK3CD expression

DNA methylation and histone acetylation are important epigenetic mechanisms that control gene expression by dictating transitions between transcriptionally active or transcriptionally silent chromatin states [50,51]. L929 fibroblasts, which express low levels of p110δ mRNA and protein, were treated with 5′-azacytidine or trichostatin A, agents known to cause DNA (hemi-)demethylation and histone hyperacetylation, respectively, creating open configurations of genomic DNA to allow binding of TFs. As a positive control, we monitored the previously documented induction in these cells of mRNA expression of the cytokines IL-6 and IFN-β by 5′-azacytidine and trichostatin A [52,53]. As can be seen from Figure 2, p110δ mRNA expression levels were not dramatically altered by any of these treatments, with a maximum increase in p110δ mRNA of around 2-fold, which was not accompanied by an induction of p110δ protein expression.

The presence of high p110δ mRNA levels is not a consequence of leukocyte-specific p110δ mRNA stability

To assess whether high expression of the p110δ protein in cells is due to increased mRNA stability, cells were treated with Actinomycin D, an inhibitor of de novo RNA synthesis, followed by measurement of mRNA decay over time. As can be seen from Figure 3A, leukocyte (A20 and EL4) and non-leukocyte (B16-BL6, 3LL and NIH-3T3) cell lines displayed very similar rates of mRNA degradation upon inhibition of mRNA synthesis, indicating that there is no difference in p110δ mRNA stability between cell types expressing high or low levels of p110δ protein. Also p110δ protein levels were not affected by Actinomycin D treatment, both in leukocytes and non-leukocytes (Figure 3B).

Identification of multiple distinct p110δ mRNA transcripts with alternate first 5′ untranslated exons

In order to identify the PIK3CD promoter, we set out to identify the transcriptional start site of the p110δ mRNA. Rapid amplification of 5′ cDNA ends (5′RACE) was used to identify the 5′UTR. BLAT alignment of the 5′RACE products led to three main observations: (1) multiple distinct p110δ transcripts exist within each cell line investigated; (2) most transcripts contain two untranslated exons (Figure 4A), which we have named exon -1 and -2 (to indicate their relative locations with respect to exon 1, which contains the putative ATG translation start site as defined in [13,54]). The -1 and -2 exons are located 11 kb and >35 kb 5′ of exon 1 in murine cells, and 19 kb and >59 kb 5′ of exon 1 in
human cells; [3] exon -1 can occur together with one of 4 identified second untranslated exons (exons -2a, -2b, -2c or -2d) in mouse cells, and with one of two -2 exons (-2a or -2b) in human cells. Thus, mouse and human PIK3CD can give rise to at least 4 and 2, respectively, distinct p110\textsubscript{d} transcripts.

Human and mouse p110\textsubscript{d} exon 1 contains an in-frame stop codon immediately upstream of the p110\textsubscript{d} start codon (for sequences, see Supporting Information, File S1), ruling out the possibility that the newly identified upstream exons are translated as part of the p110\textsubscript{d} protein. In other words, the -1 and -2 exons form the 5' UTR of the PIK3CD transcripts.

The -1 exons identified in human and mouse show a high degree of homology (data not shown) and likewise the -2a exons share a region of high homology (Figure 4B). Human exon -2a is approximately half the length of mouse exon -2a (72 bp compared to 144 bp), and its first part (1–49 bp) is homologous to the last part of mouse exon -2a (96–144 bp). Interestingly, the intronic DNA immediately upstream of human exon -2a is highly homologous with the first part of mouse exon -2a (Figure 4B). In contrast the -2b exons for human and mouse are not homologous, which may indicate that the -2b exon identified in human may not be the equivalent -2b exon identified in mouse.

Database information provides independent confirmation of several of the untranslated exons identified in this study (-1, -2a, -2b and -2c in the mouse; and -1, -2a in human), as well as additional -2 exons in the mouse (labeled with an arrowhead in Figure 4B) which we have thus far not found by 5' RACE both in

![Figure 1. Correlation between p110\textsubscript{d} protein and mRNA expression levels in murine and human cell lines.](A) Total cell lysates from the indicated cell lines were immunoblotted with antibodies to the distinct p110 isoforms or β-actin. One representative immunoblot of three independent experiments is shown. The bars represent quantification of the relative amounts of p110\textsubscript{d} protein in mouse and human cell lines, as determined in 3 independent experiments (for each cell line, the ratio of the OD of the p110\textsubscript{d} immunoblot signal was determined, relative to that of β-actin in this cell line. This value was then expressed relative to the p110\textsubscript{d}/β-actin ratio found in B16-BL6 (for the mouse lines) or HeLa (for the human lines). Values are averages of three independent experiments. (B) Quantification of p110\textsubscript{d} mRNA levels by real time RT-PCR using primers in the p110\textsubscript{d} coding region. Signals are normalised to β-actin mRNA in each cell line. Data shown are the averages of three independent experiments. doi:10.1371/journal.pone.0005145.g001)
mouse and human (>50 independent 5′RACE products sequenced; detailed data not shown).

The p110δ transcripts identified by 5′RACE always contained exon -1, and further incorporated a single -2 exon in all cases. We have found two instances where the -1/-2 exon of the p110δ mRNA configuration does not seem to occur. Firstly, during the cDNA cloning of human p110δ [13], we identified one clone (called o5) which did not contain exon -1, and which has exon -2a directly spliced onto exon 1, giving rise to a p110δ transcript that encodes full length p110δ protein (data not shown). Secondly, 2 out of 48 transcripts identified by 5′RACE in the mouse EL4 leukocyte cell line did not have -2 exons, and started with an exon -1. As shown in Table 1, all -2 and -1 exons contain a classical splice donor sequence (GT), whereas exon -1 only contains a splice acceptor sequence (AG). This is in line with the observation that each p110δ transcript identified by 5′RACE contains a single -2 exon and further indicates that the different p110δ transcripts arise individually and not from a ‘master’ p110δ transcript by alternative splicing. The presence of a splice acceptor sequence in exon -1 indicates that the p110δ transcript starting at this exon (as revealed by 5′RACE) found in mouse EL4 cells could in fact be an artefact, due to RNA degradation during the RACE experiments.

Cell type-specific usage of the multiple PIK3CD transcription start sites

We next used RT-PCR to confirm the presence of the different p110δ transcripts identified by 5′RACE and to determine which of these can be found in a panel of murine leukocyte and non-leukocyte cell lines. Forward primers, specific for each of the 5′ untranslated exons were designed, and used in combination with a common reverse primer in exon 2 (schematically shown in Figure 5A). PCR products of the predicted size were purified by agarose gel electrophoresis and verified by DNA sequencing (data not shown).

All cell lines tested (except CT26 colon carcinoma) expressed at least one type of transcript containing a -2 exon (Figure 5B). Leukocytes contained a broader variety of p110δ transcripts than non-leukocytes, with an average number of distinct p110δ transcripts of 4.3 versus 2.4 in leukocytes and non-leukocytes, respectively (summarized in bottom panel of Figure 5B).

To more accurately quantify the amount of each p110δ transcript, we next used real time RT-PCR (Figure 6). For each transcript, the PCR reaction consisted of a forward and reverse primer, which were designed to amplify a cDNA sequence of ~100 bp spanning an exon boundary specific to the transcript, and a dye-emitting probe, which bound at a sequence overlaying this exon boundary. During amplification of the cDNA sequence, cleavage of a reporter dye from the probe results in fluorescence emission, which can be directly correlated with the level of each particular p110δ transcript. For example, to measure the amount of p110δ transcripts containing exon -2a, a PCR was performed using a forward primer in exon -2a and a reverse primer in exon -
1, while the dye-emitting probe bound at this exon-exon boundary. The fluorescence detected from this PCR reaction therefore represents the amount of all transcripts specifically containing exons -2a and -1.

These experiments revealed that leukocytes express significantly higher amounts of the different p110<sub>d</sub> transcripts than non-leukocytes (Figure 6), indicating that leukocytes are likely to be more efficient at using p110<sub>d</sub> gene promoters than non-leukocytes.

In all cell lines, the transcript containing the first coding exon (exon 1; Figure 6A) was expressed at similar levels as the transcript containing the exon -2a/exon -1 boundary, which is the most abundantly expressed -2 exon (Figure 6C).

Surprisingly, the transcripts containing the exon -1/exon 1 boundary (Figure 6B) were two-fold more abundant than the transcripts containing the exon 1/exon 2 boundary (Figure 6A). This indicates that shorter but still fully processed mRNAs (i.e. with a poly A tail since oligo d(T) was used for reverse transcription) are made. These would contain a 5' UTR with at least the untranslated exon -1 and the coding exon 1, but without any of the other coding exons. Database analysis also provided evidence for such shorter p110<sub>d</sub> transcripts (marked with asterisks in Figure 4A). The mouse untranslated exon -2a was identified by a putative promoter region in/around exon -2a. Collectively, these observations indicate the presence of a putative promoter region in/around exon -2a.

**In silico analysis of PIK3CD promoter**

Alignment of the genomic sequence of flanking [and including] the 5'UTR exons of mouse PIK3CD with 8 other species revealed high homology in specific areas, indicative for functionally conserved DNA sequences, including 4 CpG islands but no TATA boxes (Figure 7A; Supporting information, File S2).

For each of the murine untranslated exons, the region spanning 500 bp upstream and 100 bp downstream of the first nucleotide was analysed for TF-binding sites and the transcription start site (TSS) prediction score within this region was assessed (Figure 7B). TF-binding sites were identified in the vicinity of all mouse untranslated exons, however a particularly condensed cluster of TF-binding sites was identified within exon -2a (Figure 7B, C). Interestingly, in human, this TF-binding cluster lies 5' of the TSS (Figure 7C; schematically shown in Figure 7D). It is unusual, but not unheard of, that promoter regions are contained within exons.

Indeed, recent work from the ENCODE project (http://www.genome.gov/10005107 and http://genome.cse.ucsc.edu/ENCODE/) has revealed that proximal TF binding sites usually fall within 1 kb of both sides, 5' and 3', of the transcription start site [60].

The TF-binding cluster of murine exon -2a was located within a CpG island (Figure 7A); was associated with a good TSS prediction score (0.9/1.0; Figure 7B) and was highly conserved across 28 species (Supporting information, File S2; Figure 7C shows the high degree homology of this region across 8 species, Figure 7D schematically shows the homology between human and mouse in this area). Collectively, these observations indicate the presence of a putative promoter region in/around exon -2a. Interestingly, 4 of the 7 different TFs identified within this binding cluster, namely ETS, IRF, NFAT and LEF, were indicated by an
asterisk in Figure 7C), have previously been associated with haematopoiesis and expression of leukocyte-specific genes (discussed in more detail below), suggesting that this TF-binding region may be involved in the high p110d expression in leukocytes.

Functional analysis of putative PIK3CD promoter elements using reporter assays

We next cloned intronic genomic DNA sequences that flank mouse exons 1, -2a and -2b (including exons -2c and -2d) at their 5’ end (Figure 7A,B; referred to as DNA fragments A–J) as well as mouse exon -2a itself (Figure 7B, DNA fragment J, which contains the TF-binding cluster), into the pGL3 reporter vector to drive expression of firefly luciferase. Vectors were transiently transfected in leukocyte and non-leukocyte cell lines and the promoter activities of the different PIK3CD DNA fragments were compared to that of the established leukocyte-specific promoter of Vav [61], and of the SV40 promoter, which is active in all cell types. The pGL3-Basic vector, which does not contain a promoter sequence upstream of firefly luciferase, was used to assess the basal level of luminescence.

The intronic genomic DNA fragments A–I did not possess significantly higher promoter activity in A20 leukocytes compared to NIH 3T3 fibroblasts (Figure 8B). This is in contrast to DNA fragment J (mouse exon-2a containing the TF-binding cluster) which had significantly higher promoter activity in the mouse macrophage cell line RAW 264.7 than in NIH 3T3 fibroblasts (Figure 8C), which was significantly higher than the leukocyte-specific Vav promoter (Figure 8C). The exon -2 fragment has also higher activity in the THP-1 monocytic cell line compared to the HEK293 (embryonic kidney) and CT26 (colon carcinoma) cell lines, again with higher activity compared to the Vav promoter (Figure 8C). Taken together with the relatively high abundance of the -2a transcripts (Figure 6C) over the other p110d exon -2 transcripts (Figure 6D–F), these data indicate that the TF binding cluster of exon -2a is the predominant promoter of p110d expression in leukocytes.

Discussion

In this study, we have explored the mechanisms by which the well-documented leukocyte-enriched expression of p110d might be

| Human | Exon | Size (bp) | Splice acceptor | 5’ end exon | 3’ end exon | Splice donor | Intron (bp) |
|-------|------|-----------|-----------------|-------------|-------------|-------------|-------------|
| -2b   | 251  | cgggggta  | GAGGCACCCA      | ACTCTGACAG  | gtaagctca   | 61,243      |
| -2a   | 59   | gcgccagc  | GCAGTCGCTC      | CGCCGGAGCG  | gtaagcgat   | 39,665      |
| -1    | 105  | ccccaacag | ATAGGGAGTC      | TTCCAGAGAG  | gtaggttgag  | 18,852      |
| 1     | 173  | catttttag | GACAACTGTC      | CATCAAGCAG  | gtaggctca   | 4,944       |
| 2     | 229  | tccctccag | CTGCTGTCGGC     | ATCGGCAAAG  | gtagctctga  |             |

Uppercase letters represent exon sequences, lowercase letters represent intron sequences.

| Murine | Exon | Size (bp) | Splice acceptor | 5’ end exon | 3’ end exon | Splice donor | Intron (bp) |
|--------|------|-----------|-----------------|-------------|-------------|-------------|-------------|
| -2d    | 150  | cttccgggc | TAGGACTTCT      | GGAGCAGTTG  | gtttattta   | 28,348      |
| -2c    | 78   | gagagaga  | ATCAGAAACC      | CTACTCAAT   | gtagattt    | 28,270      |
| -2b    | 117  | ttaggrggc | AAGAAGCAGG      | ATGTAGAAGT  | gtaagccaa   | 27,309      |
| -2a    | 144  | gttgtttt | CCTGTATATCT     | TGCTGAGCGG  | gtaaggtct   | 24,360      |
| -1     | 119  | tccctccag | ACATCTAAGG      | TACCAACAG   | gtaggttgag  | 10,759      |
| 1      | 173  | tccccacag | GAAACAGAC       | CATCAAGCAG  | gtagagcca   | 2,913       |
| 2      | 229  | tccccccag | GTGCTGTGGC      | ATCGGCAAAG  | gtaacttca   |             |

Uppercase letters represent exon sequences, lowercase letters represent intron sequences.

Table 1. Splice donor and acceptor sites in the 5’ introns/exons of PIK3CD.

| Human Exon Size | Splice donor | Human 5’ End Exon | Human 3’ End Exon | Splice acceptor | Human Intron |
|-----------------|--------------|--------------------|--------------------|-----------------|-------------|
| Human -2b       | gtaagctca    | GAGGCACCCA         | ACTCTGACAG         | cgggggta        | 61,243      |
| Human -2a       | gtaagcgat    | GCAGTCGCTC         | CGCCGGAGCG         | gcgccagc        | 39,665      |
| Human -1        | gtaggttgag   | ATAGGGAGTC         | TTCCAGAGAG         | ccccaacag       | 18,852      |
| Human 1         | gtaggctca    | GACAACTGTC         | CATCAAGCAG         | catttttag       | 4,944       |
| Human 2         | gtagctctga   | CTGCTGTCGGC        | ATCGGCAAAG         | tccctccag       |             |

Table 1 shows the splice donor and acceptor sites in the 5’ introns/exons of PIK3CD. Splice acceptor and splice donor sequences of human (top panel) and murine (lower panel) p110d exons. The untranslated exons as well as exons 1 and 2 are represented. Uppercase letters represent exon sequences, lowercase letters represent intron sequences. AG/GT splicing rule (not shown).

doi:10.1371/journal.pone.0005145.t001
achieved. We present evidence that p110\textsubscript{d} expression is mainly regulated at the transcriptional level arising from different transcripts. PIK3CD transcription appears to be constitutive and not responsive to acute cellular stimuli, at least using the agonists tested in the three cell types (NIH 3T3 and L929 fibroblasts, primary B cells and U937 myelomonocytic cells) used in our study. All cell types can express the distinct p110\textsubscript{d} mRNA transcripts but leukocytes express a greater diversity of transcripts and significantly higher amounts of the individual transcripts. In particular, the p110\textsubscript{d} transcript containing the untranslated exon -2a, was expressed at similar levels to the transcript containing coding exon 1 in a panel of murine leukocyte cell lines (compare Figure 6A with Figure 6C), indicating that the majority of p110\textsubscript{d} transcripts in a cell contain exon -2a.

We have identified a region within mouse exon -2a that contains a cluster of TF-binding sites. This TF-binding cluster is highly conserved between species and found immediately upstream of human exon -2a in human. This TF binding cluster contains at least 4 leukocyte-related TFs sites and was found to display higher promoter activity in leukocyte cell lines compared to non-leukocyte cell lines. Given that the majority of PIK3CD transcripts contain exon -2a, these data indicate that the promoter region identified within mouse exon -2a is likely to be sufficient in mediating the majority of leukocyte-specific PIK3CD gene expression.

4 of the 7 TFs identified within this binding cluster, namely ETS, IRF, NFAT and LEF have previously been associated with regulation of haematopoiesis and expression of leukocyte-specific genes. Indeed, the ETS family of TFs play important roles in the regulation of haematopoiesis [62–68]. IRF family members are highly expressed, but not exclusively, in cells of the immune system and play a pivotal role in the induction of type I IFN signalling pathways [69], proinflammatory cytokines and expression of macrophage and B cell specific genes [69,70]. NFAT family proteins are also mainly found in cells of the immune system, such as T cells, mast cells, NK cells and monocytes [71] and play a role in the regulation of various cytokines [72]. Finally, the LEF family of TFs, which are highly expressed in pre-B and T lymphocytes [73], have been linked to the regulation and expression of a number of lymphoid-specific genes [74–79].

High expression levels of p110\textsubscript{d} are also frequently observed in some non-leukocyte cancer cell lines, such as in breast carcinoma, melanoma and glioma [18]. It is possible that cancer cells upregulate or aberrantly express TFs which are, in non-cancer cells, more specific for leukocytes. It is of interest to note that a number of the TFs that bind in the exon -2a cluster have indeed been implicated in breast cancer progression, including LEF [80–83], ETS-1 [84,85], ETS-2 [86] and NFAT3 [87]. Recently, all four of these leukocyte-associated TF were identified as the most frequently differentially activated TFs in breast cancer based on a large microarray dataset [88].

We have found evidence that, among the multiple p110\textsubscript{d} transcripts, there may be mRNAs that do not encode full length p110\textsubscript{d}. Indeed, transcripts containing the exon -1/exon 1

---

**Figure 5. PIK3CD transcripts in murine cell lines and tissues, as assessed by RT-PCR** (A) Schematic representation of PCR primers used to detect distinct murine PIK3CD transcripts. In each case, a reverse primer in exon 2 and a forward primer in exon 1, -1, -2a, -2b, -2c or -2d, was used. (B) Agarose gel analysis of PCR products generated by RT-PCR (40 cycles) using primers for the different p110\textsubscript{d} mRNA transcripts in panel of murine cell lines and tissues, with the observations summarized underneath.

doi:10.1371/journal.pone.0005145.g005
boundary are more abundant than those covering the exon 1/exon 2 boundary (compare Figure 6B to Figure 6A). Current database information supports the presence of such shorter p110d transcripts (marked with an asterisk in Figure 4A). Indeed, several recent studies have reported the discovery of a new class of short promoter-associated RNA transcripts that initiate near the expected transcription start sites upstream of protein-encoding sequences [56–58] (reviewed in [55]). It remains to be seen whether these RNAs have a function, but their prevalence suggests that their synthesis may serve a functional role.

Further work is required to understand the precise mechanism of p110d gene expression. The complexity of gene regulation has been exemplified by examination of 400 protein-coding genes in 1% (30 million bases) of the human genome as part of the ENCODE project [60], which revealed that 80% of these genes had additional exons, many of which were located thousands of bases away from the coding exons. Also many novel transcription start sites were found, many located thousands of bases away from the known start sites, while 25% of the promoters discovered were at the 3' end of the genes rather then at the 5' end. It is therefore highly likely that p110d expression will be subject to additional levels of control rather than by simple proximal promoter elements.

The data presented are the first to shed light onto the leukocyte-enriched expression of PIK3CD. Further investigations are needed to identify which TF-binding sites are critical in driving PIK3CD gene expression and whether cells of non-leukocyte origin, such as breast cancer cells, are able to utilize this putative promoter.

Interference with PIK3CD expression at the promoter level may offer a novel therapeutic target in cases of aberrant p110d overexpression, as observed in some cancers [18].

**Materials and Methods**

**Antibodies and reagents**

Antibodies to class IA PI3Ks were generated in-house or purchased from Santa Cruz Biotechnology (p110d, sc-602). Cell culture reagents were purchased from Invitrogen, recombinant mouse TNF was provided by Peter Brouckaert (Ghent University, Belgium), other reagents were from Sigma: Actinomycin D (856258), 5'-azacytidine (A2358), trichostatin A (T8552), antibodies to β-actin (A5441).

**RNA extraction, 5’Rapid Amplification of cDNA Ends (5'RACE), Reverse Transcription (RT)-Polymerase Chain Reaction (PCR) and Real Time RT-PCR**

Total RNA was extracted from cells using the RNaseasy mini kit (Qiagen, 74104). mRNA was subsequently reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen, 18064) and oligo d(T) primers, and subsequently used in 5'RACE, RT-PCR or real time PCR.

For 5'RACE, the FirstChoice® RLM-RACE kit (Ambion, 1700) was used following the manufacturer's protocol using the following outer primers: murine 5'-CAGATCAGCTCCTCA-TTGGCACT-3', human 5'-GTTTTCACCGTGCCTG-3' and inner primers: murine 5'-ACTTGAACTCACC-3'.
Figure 7. Bioinformatic analysis of potential promoter elements and TF binding sites in PIK3CD. (A) Schematic representation of the murine p110\(\delta\) 5'UTR. The upper panel shows the five untranslated murine p110\(\delta\) exons (and exon 1) in their genomic context, with below (in descending order): the mouse RefSeq genes, CpG islands, homology with 9 species (rat, human, chimp, rhesus, dog, cow, armadillo, elephant, tenrec, plotted against the murine sequence) and the genomic fragments that were subcloned for use in gene reporter assays. (B) Locations of the conserved TF binding sites in the 600 bases (500 upstream to 100 downstream of the transcript start site) in the forward strand flanking the different 5' untranslated exons of mouse p110\(\delta\) gene. The exon start sites are indicated by the vertical arrows, the TF binding sites found on the forward strand are shown as blue boxes above the TSS score graphs. Also shown is the degree of cross species (28 species) genomic conservation as calculated by the phastCons program \[94\] from a minimum of 0.0 to a maximum of 1.0. The genomic DNA fragments subcloned into the PGL3 reporter vector are shown underneath. (C) Alignment and conservation of the TF binding cluster identified in mouse exon -2a with genomic sequences upstream of the translation start site of PIK3CD of 7 other species. (D) Schematic representation of TF binding cluster location in relation to exon -2 in human and mouse.

doi:10.1371/journal.pone.0005145.g007

For real time RT-PCR amplification TaqMan Universal PCR Mastermix (4304437) and primer mixes containing a FAM reporter probe (TaqMan Gene Expression Assay) were obtained from Applied Biosystems. SYBR Green (Qiagen, 204143) was used for quantifying 18S RNA. Exon-specific primer sets and probes were designed to identify transcripts containing exon -1, -2a, -2c and -2d. For exon -2a the following primer sequences were used; forward primer 5'-TCGCGCCTAGCCTTGG-3', reverse primer 5'-GGCATCAGCGGGCTTCA-3' and FAM reporter sequence 5'-CTCAGCTCCTTAGATGTCGGTC-3'. For exon -2b the following primer sequences were used; forward primer 5'-AGTGTCGTGTCGCTGACTTCAAGAA' and FAM reporter sequence 5'-CAGCTCCTTATAGTGTCTCAGTGTC-3' and FAM reporter sequence 5'-CAGCTCCTTATAGTGTCTCAGTGTC-3'. For each transcript of interest, known amounts of plasmids with this transcript were used to create a standard curve. Real-time PCR generated a series of C\(_T\) values (the PCR cycle at which amplification of each target gene is first detected) for endogenous and plasmid-born cDNA, which allowed for
the determination of mRNA copy numbers for each individual gene.

**Western blot**

Cells were lysed and immunoblotted for PI3K expression as described before [89]. Primary antibodies were detected using fluorescently-labeled species-specific secondary antibodies (anti-mouse IRDye 800-conjugated (Rockland) and anti-rabbit AlexaFluor 680-conjugated (Molecular Probes). Quantification was done using an Odyssey infrared scanner (LICOR) using the manufacturer's software. Signal intensities were normalized for an internal loading control such as β-actin or GAPDH.

**Bioinformatic analysis of putative promoter elements in PIK3CD genes**

The upstream sequences for the five untranslated exons of the murine p110d gene were inspected within the February 2006 (NCBI build 36) assembly of the Mouse genome using the UCSC genome browser [90]. Regions spanning 500 bp upstream and 100 bp downstream of the first nucleotide of each exon were analysed. The corresponding multiple species alignment was extracted using the Vertebrate Multiz Alignment & Conservation track [91] within the UCSC genome browser. The alignments were then screened for conserved TF binding sites using MatInspector [92] and a vertebrate factors subset of a proprietary database of Genomatix. In addition the candidate regions were inspected with Eponine [93], a probabilistic method for detecting transcription start sites, using a threshold of 0.9.

**Reporter gene assays**

PCR amplification of genomic DNA from C57Bl/6 mice was used to generate fragments for the reporter assays. The amplified PCR products were inserted into the pGL3 reporter vector (Promega). Transfections of NIH3T3 and A20 cells were performed using Qiagen Superfect or electroporation, respectively. Equal number of cells were washed and lysed, using Promega lysis buffer (to normalize for transfection efficiency) and then assayed for luciferase activity using the firefly luciferase substrate from Promega on the MicroBeta workstation (Perkin Elmer). The luciferase activity was normalized using a luciferase gene in a pGL3 reporter vector under the control of the SV40 promoter as well as a promoterless luciferase/pGL3 reporter vector. DNA of the lymphocyte-specific Vav promoter (construct HS21 mentioned in Ref. [61], which we cloned from the original β-galactosidase.
reporter construct into a pGL2 luciferase reporter vector) was used as a positive control.

Supporting Information

File S1  5’ RACE product sequences Sequences of the different murine (Mm) and human (Hs) p110α transcripts as identified by 5’ RACE. Found at: doi:10.1371/journal.pone.0005145.s001 (0.03 MB DOC)

File S2  Annotated multiple species alignments. This file shows the conserved TF binding sites for the different murine and human p110α mRNA transcripts as identified by 5’ RACE. Found at: doi:10.1371/journal.pone.0005145.s002 (0.19 MB DOC)

References

1. Vanhaesebroeck B, Leevens SJ, Ahnadi K, Timm J, Katso R, et al. (2001) Synthesis and function of 3-phosphorylated inositol lipids. Annu Rev Biochem 70: 533–602.
2. Vanhaesebroeck B, Leevens SJ, Panayotou G, Waterfield MD (1997) Phosphoinositide 3-kinases: a conserved family of signal transducers. Trends Biochem Sci 22: 267–272.
3. Krugmann S, Hawkins PT, Pryor N, Brashelman S (1999) Characterizing the interactions between the Two Subunits of the p110α/p110gamma Phosphoinositide 3-Kinase and Their Role in the Activation of This Enzyme by Gbeta gamma Subunits. J Biol Chem 274: 17132–17138.
4. Stephens LR, Equiambis H, Lui M, Cooke F, et al. (1997) The Gbeta[gamma] Sensitivity of a PI3K Is Dependent upon a Tightly Associated Adaptor, p101. Cell 89: 105–114.
5. Surer S, Casadewj J, Ferguson GJ, Davidson K, Hawkins P, et al. (2003) p84, a New Gbeta[gamma]-Activated Regulatory Subunit of the Type IB Phosphoinositide 3-Kinase p110α. Science 297: 9927–9936.
6. Guillermet-Gibert J, Bjorkdol K, Salpekar A, Gonnella C, Ramadani F, et al. (2008) The p110beta isoform of phosphoinositide 3-kinase signals downstream of G protein-coupled receptors and is functionally redundant with p101gama. Proc Natl Acad Sci USA 105: 8292–8297.
7. Jia S, Liu Z, Zhang S, Liu P, Zhang L, et al. (2008) Essential roles of PI3K/p110beta in cell growth, metabolism and tumorigenesis. Nature 454: 776–779.
8. Cirado E, Iezzi M, Marone R, Marrone S, Currino C, et al. (2008) Phosphoinositide 3-Kinase p110bet[alpha] Activity: Key Role in Metabolism and Mammary Gland Cancer But Not Development. Sci Signal 1: ra3.
9. Kok K, Geering B, Vanhaesebroeck B (2006) Regulation of phosphoinositide 3-kinase in health and disease. Trends Biochem Sci Accepted for publication.
10. Li B, Okabe I, Bernard DJ, Wynshaw-Boris A, Nussbaum RL (1999) P110delta, a novel phosphoinositide 3-kinase in leukocytes. Proc Natl Acad Sci USA 105: 8292–8297.
11. Geering B, Guertler PR, Noek G, Gharbi SI, Vanhaesebroeck B (2007) Class IA phosphoinositide 3-kinases are obligate p85-p110 heterodimers. Proc Natl Acad Sci USA 104: 7090–7094.
12. Vanhaesebroeck B, Wehmij MH, Kotani K, Stein R, Warne PH, et al. (1997) PI110delta, a novel phosphoinositide 3-kinase in leukocytes. Proc Natl Acad Sci USA 104: 7090–7094.
13. Geering B, Guertler PR, Noek G, Gharbi SI, Vanhaesebroeck B (2007) Class IA phosphoinositide 3-kinases are obligate p85-p110 heterodimers. Proc Natl Acad Sci USA 104: 7090–7094.
14. Bi L, Okabe I, Bernard DJ, Nussbaum RL (1999) Proliferative defect and embryonic lethality in mice homozygous for a deletion in the p110α subunit of phosphoinositide 3-kinase. J Biol Chem 274: 10963–10968.
15. Patrucco E, Notte A, Barberis L, Selvettella G, Maffei A, et al. (2004) PI3Kgamma modulates the cardiac response to chronic pressure overload by distinct kinase-dependent and -independent effects. Cell 118: 373–387.
16. Bi L, Okabe I, Bernard DJ, Nussbaum RL (2002) Early embryonic lethality in mice deficient in the p110beta catalytic subunit of PI 3-kinase. Mamm Genome 13: 7677–7688.
17. Hu P, Mondino A, Skolnik EY, Schlessinger J (1993) Cloning of a novel, ubiquitously expressed human phosphatidylinositol 3-kinase and identification of its binding site on p85. Mol Cell Biol 13: 7677–7688.
18. Chantry D, Vojtek A, Kashishan A, Holtzman DA, Wood C, et al. (1997) p110delta, a novel phosphatidylinositol 3-kinase catalytic subunit that associates with p85 and is expressed predominantly in leukocytes. J Biol Chem 272: 19296–19291.
19. Eckholt BJ, Ahmed AI, Davies M, Papakonstanti EA, Pearce W, et al. (2007) Control of Axonal Growth and Regeneration of Sensory Neurons by the p110delta PI 3-Kinase. PLoS ONE 2: e696.
20. Sawyer C, Sturje J, Bennett DC, O'Hare MJ, Allen WE, et al. (2003) Regulation of breast cancer cell chemotaxis by the phosphoinositide 3-kinase p110delta. Cancer Res 63: 1667–1675.
21. Hirsch E, Wyman MP, Patrucco E, Tolosano E, Bagarelli-Leva G, et al. (2000) Analysis of the murine phosphoinositide 3-kinase gamma gene. Gene 256: 69–81.
36. Scott EW, Fisher RC, Olson MC, Kehrl EW, Simon MC, et al. (1997) PI3 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. Immunity 4: 373–47.

37. Su GH, Chen HM, Muthusamy N, Garrett-Sinha LA, Baezoch D, et al. (1997) Defective B cell receptor-mediated responses in mice lacking the Ets protein, Spi-B. EMBO J 16: 7118–7129.

38. Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, et al. (1997) Yolk sac agnostic defect and intra-embryonic apoptosis in mice lacking the Ets-related transcription factor Fli-1. Cell 90: 1457–1473.

39. Ozato K, Tailor P, Kubota T (2007) The interferon regulatory factor family in host defense: mechanism of action. J Biol Chem 282: 20063–20069.

40. Paun A, Pitha PM (2007) The IRF family, revisited. Biochimie 89: 744–753.

41. Kao A (1994) NF-ATp: a transcription factor required for the coordinated induction of several cytokine genes. Immunity 1: 117–125.

42. De Boer ML, Mordvinov VA, Thomas MA, Sanderson CJ (1999) Role of nuclear factor of activated T cells (NFAT) in the expression of interleukin-5 and other cytokines involved in the regulation of hematopoietic cells. Int J Biochem Cell Biol 31: 1229–1236.

43. Travis A, Amstadler A, Belanger C, Grosseck R (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. Genes Dev 5: 890–894.

44. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tontogawa S, et al. (1992) RAG-1-deficient mice have no mature B and T lymphocytes. Cell 68: 869–877.

45. Shinkai Y, Rathbun G, Lam KP, Olte EM, Stewart V, et al. (1992) RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate VDJ recombination. Cell 68: 879–887.

46. Jin Zx, Kishi H, Wei Xc, Matsuda T, Saito S, et al. (2002) Lymphoid enhancer-binding factor-1 binds and activates the recombination-activating gene-2 promoter together with c-Myc and Pax-5 in immature B cells. J Immunol 169: 3783–3792.

47. Brickett AG, Gossage DL, Dusing MR, Wighton DA (1995) Identification of a murine homolog of the human adenosine deaminase thymic enhancer. Gene 167: 261–266.

48. Sawada S, Litman DR (1991) Identification and characterization of a T-cell enhancer adjacent to the murine CD4 gene. Mol Cell Biol 11: 5306–5315.

49. Leiden JM, Thompson CB (1994) Transcriptional regulation of T-cell genes during T-cell development. Curr Opin Immunol 6: 231–237.

50. Ayayan A, Civenni G, Ciaroni L, Morel C, Muller N, et al. (2006) Increased Wnt signaling triggers oncogenic conversion of human breast epithelial cells by a Notch-dependent mechanism. Proc Natl Acad Sci U S A 103: 3799–3804.

51. Geiman TM, Robertson KD (2002) Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together? J Cell Biochem 87: 723–730.

52. Vanden Berghe W, De Bosscher K, Boone E, Plaisance S, Haegeman G (1999) The nuclear factor-kappaB engages CBP/p300 and histone acetyltransferase activity for transcriptional activation of the interleukin-6 gene promoter. J Biol Chem 274: 12091–12096.

53. Shinkai Y, Rathbun G, Lam KP, Oltz EM, Stewart V, et al. (1992) RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate VDJ recombination. Cell 68: 879–887.

54. Clayton E, McAdam S, Coadwell J, Chantry D, Turner M (2001) Structural organization of the mouse phosphatidylinositol 3-kinase p110d gene. Biochem Biophys Res Commun 280: 1328–1332.

55. Span PN, Manders EJ, Heuvel JJ, Thomas CM, Bosch RR, et al. (2002) Expression of the transcription factor Ets-1 is an independent prognostic marker for relapse-free survival in breast cancer. Oncogene 21: 6506–6509.

56. Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322: 1845–1848.

57. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW (2000) The antisense transcripts of human genes. Cell 92: 1855–1857.

58. Preker P, Nielen J, Kammerl S, Lykke-Andersen S, Christensen MS, et al. (2001) 3′-Regulatory region of the murine CD4 gene. Mol Cell Biol 21: 8506–8515.