



Gómez-Castañeda, Eduardo (2018) *Uncovering the BCR-ABL1 tyrosine kinase independent signature in chronic myeloid leukaemia stem cells*. PhD thesis.

<https://theses.gla.ac.uk/39043/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Uncovering the BCR-ABL1 tyrosine kinase independent signature in chronic myeloid leukaemia stem cells

Eduardo Gómez Castañeda

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

Institute of Cancer Sciences

College of Medical, Veterinary & Life Sciences

University of Glasgow



University
of Glasgow

September 2018

Abstract

Chronic myeloid leukaemia (CML) is characterised by the presence of the fusion protein BCR-ABL1. The addiction of CML cells to the tyrosine kinase (TK) activity of the oncoprotein has been successfully exploited by the introduction of tyrosine kinase inhibitors (TKI), such as imatinib, which have shown a great success at managing the disease. However, these compounds fail to eradicate a primitive cell population, the leukaemic stem cells (LSCs), which persist in the patients. This translates in the need of life-long therapy for most of the patients, meaning a higher risk of treatment side effects and the prolonged psychological burden of living a leukaemia patient. Life-long therapy is also translating in a continuous increase in CML prevalence in developed countries and sustaining a big patient population on TKI treatment is becoming a challenge for national health systems.

Recent reports in chronic myeloid leukaemia biology have confirmed that CML LSCs are not addicted to the TK activity of BCR-ABL1 and they retain repopulation and leukaemic properties even during BCR-ABL1 TK inhibition. Thus, the discovery of new therapeutic targets capable of eliminating this cell population is required for curing the disease. Previous reports have already shown great success at reducing the number of CML LSCs by targeting JAK2, STAT5, EZH2, MYC and p53 pathways as well as autophagocytosis. However, none of them have shown complete eradication of the clone and they failed to define a global gene expression signature that may explain the persistence of CML LSCs during TKI treatment.

In this thesis, microarray analyses revealed (I) 527 consistently de-regulated genes in CML LSCs compared with normal HSCs and (II) 5,706 genes not affected by TKI treatment in CML CD34⁺ cells. The comparison of both lists revealed a 60 genes signature that is differentially expressed in CML LSC (as compared with normal HSC) and not affected by TKI treatment of CML CD34⁺ cells. Of the 60, 4 genes (*CD33*, *CHST2*, *PPIF* and *ERG*) were validated in a set of independent patients and controls by qPCR. *CD33*, a myeloid cell surface marker, was found to be upregulated in CML and can be targeted by the clinical grade compound gemtuzumab-ozogamicin (GO). GO proved effective at targeting CML CD34⁺ cells *in vitro* and after 72 hours of treatment it had an IC₅₀ of 136ng/mL, which was 19 times lower than the BCR-ABL1⁻ control and about 7 times lower than the FDA-approved dose for acute myeloid leukaemia (≈1000ng/mL). Importantly, the combination of GO with 2μM imatinib presented a mainly additive effect and the combination treatment had an IC₅₀ of 195ng/mL using the imatinib alone treatment as baseline and the number of colony forming cells was reduced in a concentration dependent

manner. Additionally, an increase in γ H2AX, a marker for double strand breaks in the DNA, was observed in a concentration dependent manner, consistent with the mechanism of action of GO, which induces DNA damage. Interestingly, GO induced cell-cycle entry in the CML cells in a concentration dependent manner, opposing the antiproliferative effect of imatinib. Similar effects were observed when (I) the cells were initially treated 72 hours with imatinib and then 72 hours with GO and (II) when the cells were initially treated 72 hours with GO and then 72 hours with imatinib. The sequential treatments revealed that GO is effective at targeting the cells that remained after TKI treatment and that a concentration of at least 100ng/mL of GO is needed to observe medium-term effects in CML CD34⁺ cells when they are allowed to recover. Global gene expression differences after GO treatment were enrichment in homeostasis and inflammation pathways, which is consistent with the increase in cell death and apoptosis and may represent an activation of repopulating pathways in CML LSCs.

Another clinical need in CML is the development of molecular biomarkers that complement the current EUTOS score. Although the EUTOS score has a high specificity, which provides a great confidence in the high-risk patients, it only has 16% sensitivity for predicting the patients who will not achieve progression free survival. Therefore, a high proportion of patients with bad prognosis that are currently scored as low-risk would benefit from the use of a more sensitive score. With this in mind, the TKIi signature was tested as a biomarker for TKI response and disease aggressiveness. The TKIi signature was a better classifier than both (I) all the RefSeq transcripts analysed in the microarray and (II) 97.1% of randomly generated probe sets of the same size, showing its potential for predicting response to TKI. However, the TKIi signature was not found informative in predicting disease aggressiveness.

Interestingly, most of the genes present in the TKIi signature are deregulated in the same direction compared with normal controls across all the phases of CML. This suggests that any treatment targeting the signature has the potential to be effective across all the phases of the disease, which may accelerate its use in the clinical practice due to the lack of therapeutic options for treating CML blast crisis.

Taken together, the work presented in this thesis confirms the existence of a BCR-ABL1 transcriptional signature in CML LSCs. Also, it shows that targeting CD33, a member of the TKIi signature, reduces the number of CML CD34⁺ cells and induces a transcriptional and phenotypic change towards a cycling and repopulating cell population. Additionally,

the use of the TKI signature has shown potential as a molecular biomarker for predicting TKI response in CML patients.

Table of contents

Abstract	2
Table of contents	5
List of tables	11
List of figures	12
Acknowledgements	16
Related publications	18
Author’s declaration	19
List of abbreviations	20
1 Introduction	21
1.1 Haematopoiesis	21
1.2 Cancer	29
1.2.1 Chronic Myeloid Leukaemia (CML)	30
1.2.1.1 Clinical presentation of CML	30
1.2.1.2 Molecular and cellular characteristics of CML	31
1.2.1.3 Treatment of CML	33
1.2.1.4 Prognosis scores in CML	34
1.2.1.5 CML LSC persistence and evidence of a BCR-ABL1 TK independent signature	36
1.3 Aims	40
2 Materials and methods	41
2.1 Materials	41
2.1.1 Primary patient material	41
2.1.2 Tissue culture solutions	41
2.1.2.1 RPMI+	41
2.1.2.2 RPMI++	41
2.1.2.3 DMEM+	41
2.1.2.4 Serum free media (SFM)	43
2.1.2.5 Physiological growth factors (PGF)	43
2.1.2.6 FACS solution	43
2.1.2.7 DAMP	43
2.1.2.8 Magnetic activated cell sorting buffer	43

2.1.2.9	Gemtuzumab-ozogamicin (GO)	43
2.1.2.10	Imatinib (IM)	43
2.1.3	Flow cytometry	44
2.1.3.1	Antibodies	44
2.1.4	Molecular biology	44
2.2	Methods	45
2.2.1	Cell biology	45
2.2.1.1	Cell counts	45
2.2.1.2	Recovery of frozen primary cells	45
2.2.1.3	Magnetic activated cell sorting (MACS)	45
2.2.1.4	Surface antigens staining and detection	46
2.2.1.5	Culture of CML CD34 ⁺ cells for validation of the TKIi genes	46
2.2.1.6	Culture of primary cells for drug response studies	46
2.2.1.7	Assessment of apoptosis	47
2.2.1.8	Staining and detection of intracellular antigens	48
2.2.1.9	Colony forming cell (CFC) assay	48
2.2.1.10	Long-term culture	48
2.2.1.10.1	Preparation of the feeder cells layer	48
2.2.1.10.2	Seeding and culture of primary cells on feeder cells	49
2.2.1.10.3	Harvest	49
2.2.1.11	Cell lines drug response analysis	50
2.2.1.12	Resazurin	50
2.2.2	Molecular biology	50
2.2.2.1	RNA extraction	50
2.2.2.2	Primer design	50
2.2.2.3	Retrotranscription	51
2.2.2.4	Quantitative polymerase chain reaction (qPCR)	51
2.2.2.5	Fluidigm	51
2.2.2.6	RNA sequencing	52
2.2.3	Bioinformatics	53
2.2.3.1	Microarray analysis	53
2.2.3.2	<i>In silico</i> validation of gene lists	54
2.2.3.3	Principal components analysis (PCA)	55
2.2.3.4	Generation of a classifier: Support vector machine (SVM)	55
2.2.3.5	qPCR analysis	55

2.2.3.6	RNAseq analysis.....	56
2.2.3.7	Pathway overrepresentation analysis.....	56
2.2.3.8	Drugs synergy calculation.....	57
3	Results (I): Identification of a TKI independent signature in CML LSCs.....	58
3.1	Introduction	58
3.2	Identification of the TKIi signature using transcriptomic datasets	60
3.2.1	Description of the microarray datasets.....	60
3.2.1.1	Dataset TKIFP: CML LSCs' transcriptional response to TKI.....	60
3.2.1.2	Dataset CMLDV: transcriptional differences between CML and normal	60
3.2.1.3	Dataset CMLMC: transcriptional differences between CML and normal	61
3.2.2	Definition the threshold for no-change in microarray.....	63
3.2.3	Differential gene expression and non-changing genes in the microarray datasets	66
3.3	Validation of the TKIi signature	69
3.3.1	In silico validation of the TKIi signature	69
3.3.2	In vitro validation of TKIi signature: primer optimization	70
3.3.3	Validation of the TKIi signature by qPCR in CD34 ⁺ cells	72
3.4	Discussion	78
4	Results (II): Investigation of the potential role of the TKI independent signature as a biomarker in CML LSCs.....	82
4.1	Introduction	82
4.2	The TKIi signature is able to predict TKI response in CML	84
4.2.1	Description of the TKI response microarray dataset.....	84
4.2.2	Replication of the analysis conditions of RNRMW failed to return the same results than presented in the publication	84
4.2.3	Failing to correct for false discovery rate reproduce similar results to the analysis presented in the original publication	89
4.2.4	The overlap between the TKIi signature and the genes differentially expressed between TKI responders and non-responders is bigger than the overlap expected by chance	89

4.2.5	The TKIi signature can discriminate between TKI responders and non-responders	90
4.3	The TKIi signature is not able to predict disease aggressiveness	99
4.3.1	Description of the disease aggressiveness microarray dataset.....	99
4.3.2	Differential expression analysis revealed no significant differences between aggressive and indolent patients.....	99
4.3.3	The TKIi signature is not a marker for disease aggressiveness	99
4.4	Disease progression is associated with the expression of the TKIi signature	105
4.4.1	Disease phase and cell population are strongly associated with the variance in the dataset.....	105
4.4.2	Differential gene expression reveals an enrichment on TKIi genes in BC and AP	105
4.4.3	Selecting only the TKIi genes increases the discrimination between CML phases	106
4.5	Discussion	111
5	Results (III): CD33 as a therapeutic target in CML CD34 ⁺ cells	114
5.1	Introduction	114
5.2	Results	116
5.2.1	K562 cell line expresses CD33 and is sensitive to treatment with GO.....	116
5.2.2	Treatment with GO increases cell size in K562 cells	116
5.2.3	Treatment with GO has a modest impact on apoptosis in K562 cells	116
5.2.4	GO and IM have additive effect when combined in K562 or BV173 cell lines	116
5.2.5	Binding of GO to the cell surface is dose dependent.....	117
5.2.6	GO induces DNA damage in a dose dependent manner and it is independent of IM treatment	117
5.2.7	CML CD34 ⁺ cells express CD33 on the cell surface at similar levels to nCML CD34 ⁺ cells.....	123
5.2.8	Binding of GO to the cell is dependent on the levels of CD33 on the cell surface	123
5.2.9	IM treatment does not affect CD33 levels on the surface of the cells	123

5.2.10	GO reduces cell number in CML CD34 ⁺ cells while having a minor effect on nCML CD34 ⁺ cells after 72h of treatment	123
5.2.11	GO induces apoptosis in CML CD34 ⁺ cells after 72h of treatment.....	125
5.2.12	GO eliminates colony forming cells (CFCs) in both CML and nCML after 72h of treatment	128
5.2.13	Detection of LTC-IC after 72h of GO treatment is patient dependent and no differences were found between treatments	129
5.2.14	GO increases γ H2AX levels in CML CD34 ⁺ cells but not in nCML	135
5.2.15	GO promotes cell cycle entry in CML cells while increasing the proportion of quiescent cells in nCML after 72h of treatment	135
5.2.16	GO reduces the number of CML CD34 ⁺ cells after 72h of culture followed by 72h of treatment and its effect is additive to a 72h pre-treatment with IM.....	136
5.2.17	Pre-treatment with IM protects CML cells from entering apoptosis	139
5.2.18	IM pre-treatment increases CFC concentration while reducing its total number in CML.....	139
5.2.19	No differences in the number of LTC-ICs were found in GO treated cells previously treated with IM 2 μ M for 72h	141
5.2.20	Pre-treatment with IM does not affect the level of γ H2Ax, which increases with increasing concentrations of GO.....	141
5.2.21	A delayed start on GO treatment induces increasing percentage of nCML cells entering S/G ₂ /M phases of the cell cycle	142
5.2.22	Treatment with IM after treating CD34 ⁺ cells with GO synergises with GO in reducing cell number.....	151
5.2.23	Medium-term effect of GO in percentage of viability can be observed in concentrations equal of higher than 100ng/mL.....	151
5.2.24	IM treatment after GO treatment had a moderate effect at reducing the total number of CFCs in CML	152
5.2.25	IM treatment after GO treatment further reduces the number of LTC-ICs in CML	152
5.2.26	Treating CML cells with IM that were previously treated with GO increases the levels of γ H2AX.....	152

5.2.27	Treatment with GO reduced the percentage of cells in G ₀ 72h after stopping the treatment in both CML and nCML and seemed to be independent of IM	153
5.2.28	Understanding the mechanism of action of GO by investigating whole transcriptome expression changes by RNAseq after 72h of treatment	162
5.3	Discussion	172
6	General discussion	176
6.1	New therapeutic approaches are necessary for curing CML	176
6.2	Biomarkers for TKI response and prognosis	181
7	Appendix I	184
8	Appendix II	185
9	Appendix III	192
10	Bibliography	201

List of tables

Table 2-1. Summary of the CML primary patient samples.	42
Table 2-2. Summary of the nCML patient samples..	42
Table 3-1. Gene expression changes of the TKIi genes with successful amplification in Fluidigm.	75
Table 4-1. Over and under-represented signalling pathways in the DE genes between AP and normal controls.....	108
Table 4-2. Overrepresented signalling pathways in the DE genes between BC and CP....	108
Table 4-3. TKIi genes are overrepresented in the DE genes of all CML phases.....	108
Table 5-1. Top 10 overrepresented pathways after 72h of treatment with GO.....	169
Table 5-2. Top 10 overrepresented GO-terms after 72h of treatment with GO.....	169
Table 5-3. Overrepresented pathways after 72h of treatment with IM.	169
Table 5-4. Top 10 overrepresented GO-terms after 72h of treatment with IM.....	170
Table 5-5. Top 10 overrepresented pathways after GO+IM combination treatment compared with the NDC.....	170
Table 5-6. Top 10 overrepresented GO-terms after GO+IM combination treatment compared with the NDC.....	170
Table 5-7. Top 10 overrepresented pathways after GO+IM combination treatment compared with IM treatment.....	171
Table 5-8. Top 10 overrepresented GO-terms after GO+IM combination treatment compared with IM treatment.....	171
Table 7-1. Primer sequences of the primers used for qPCR	184
Table 8-1. GO vs NDC (pathways).....	185
Table 8-2. IM vs NDC (pathways).....	186
Table 8-3. Combination treatment (GO+IM) vs NDC (pathways).....	186
Table 8-4. Combination treatment (GO+IM) vs IM (pathways).	188
Table 9-1. GO vs NDC (GO-terms).....	192
Table 9-2. IM vs NDC (GO-terms).....	194
Table 9-3. Combination (GO+IM) vs NDC (GO-terms).	195
Table 9-4. Combination treatment (GO+IM) vs IM. (GO-terms).....	198

List of figures

Figure 1-1. Different haematopoiesis models.	25
Figure 2-1. Classification of the cells in viable, apoptotic and necrotic by AnnexinV and DAPI staining.	47
Figure 3-1. Summary of the microarray analysis performed in this chapter for defining the TKI independent (TKIi) transcriptomic signature.	59
Figure 3-2. The different groups analysed in each dataset form distinct clusters in the PCA.	62
Figure 3-3. Explanation of what an equivalence test	64
Figure 3-4. The noise interval was set between -0.5 and +0.5.	65
Figure 3-5. The 4 olfactory receptor (OR) genes present in the TKIi list cluster together in the genome at 1q44.	67
Figure 3-6. A total of 60 genes were identified as TKIi.	68
Figure 3-7. The TKIi signature is unlikely to be generated by chance.	71
Figure 3-8. Primer efficiencies in K562 cells.	72
Figure 3-9. The noise interval in the 48.48 Fluidigm chip.	73
Figure 3-10. A list of 6 genes was found to be TKIi using nCMLp samples as control.	76
Figure 3-11. A list of 4 genes was found to be TKIi in using normal samples as control. .	77
Figure 4-1. TKI response is not associated with any of the first 8 PCs in the training set.	85
Figure 4-2. TKI response is not associated with any of the first 7 PCs in the validation set.	86
Figure 4-3. Correction for FDR reduces the number of differentially expressed probe sets in the validation set to only <i>HOXA1</i>	87
Figure 4-4. The overlap of the differentially expressed RefSeq transcripts is unlikely to happen by chance.	88
Figure 4-5. Selecting only the TKIi RefSeqs increases the discrimination between responders and non-responders in the training set.	92
Figure 4-6. Selecting only the TKIi RefSeqs increases the discrimination between responders and non-responders in the validation set.	93
Figure 4-7. Selecting only for the probe sets differentially expressed in the training set increases the discrimination between responders and non-responders in the training set. .	94
Figure 4-8. Selecting only for the probe sets differentially expressed in the training set increases the discrimination between responders and non-responders in the validation set.	95

Figure 4-9. Selecting only for the probe sets present in McWeeney's classifier has the largest effect in discriminating between responders and non-responders in the training set..	96
Figure 4-10. Selecting only for the probe sets present in McWeeney's classifier has the largest effect in discriminating between responders and non-responders in the validation set.	97
Figure 4-11. The TKIi signature and the McWeeney classifier have higher AUCs than random.....	98
Figure 4-12. AIAY's PCA reveals association of gender, aggressiveness and age with PCs 2, 3 and 4	101
Figure 4-13. Differential gene expression analysis revealed no significantly DE genes between aggressive and indolent disease in CML	102
Figure 4-14. Selecting only for the TKIi transcripts reduced the association of the PCs with any of the known factors (disease aggressiveness, age and gender).....	103
Figure 4-15. Selecting only for the McWeeney classifier transcripts associated the PC1 with disease aggressiveness and age	104
Figure 4-16. Principal components are strongly associated with all the three known factors in CMLMC: cell population, phase of the disease and presence of BCR-ABL1	107
Figure 4-17. Selection for the TKIi genes increases the variance explained by the principal component 1 and the association of disease phase with the principal components	109
Figure 4-18. The direction of relative expression of the TKIi genes is constant in all phases of CML when compared with normal controls.	110
Figure 5-1. K562, a CD33-expressing cell line, has a lower IC ₅₀ than BV173, a non CD33-expressing cell line.	118
Figure 5-2. GO increases size of K562 cell line and induces apoptosis.	119
Figure 5-3. Relative cell concentrations to NDC based on resazurin absorbance.	120
Figure 5-4. Bliss coefficients for each IM-GO combination and binding of GO to the cell surface.	121
Figure 5-5. The level of γ H2AX increases in response to GO but it is not affected by IM in K562 cells	122
Figure 5-6. Expression levels of CD33 on the cell surface were the same between CML and nCML and are not affected by IM.....	126
Figure 5-7. Treatment with GO reduces CML cell number after 72h of treatment.....	127
Figure 5-8. GO reduced the percentage of viable cells in both CML and nCML.....	130
Figure 5-9. Colony counts after 72h of treatment with GO $\pm$ IM.	131
Figure 5-10. CFC concentration after 72h of treatment with GO $\pm$ IM.....	132

Figure 5-11. Colony counts after 72h of treatment with GO±IM and 6 weeks on culture with stroma cells	133
Figure 5-12. CFC concentration after 72h of treatment with GO±IM and 6 weeks on culture with stroma cells.	134
Figure 5-13. Levels of γ H2AX increases in CML after 72h of treatment with GO..	137
Figure 5-14. Differences on the percentage of cells in each cell cycle phase compared with the NDC.	138
Figure 5-15. Treatment with GO after IM reduces CML cell number.....	143
Figure 5-16. GO after IM/no treatment reduced the percentage of viable cells in both CML and nCML	144
Figure 5-17. Colony counts after sequential treatment IM/no drug → GO.....	145
Figure 5-18. CFC concentration after sequential treatment IM/no drug → GO.....	146
Figure 5-19. Colony counts after sequential treatment IM/no drug → GO and 6 weeks on culture with stroma cells	147
Figure 5-20. CFC concentration after sequential treatment IM/no drug → GO and 6 weeks on culture with stroma cells.	148
Figure 5-21. Levels of γ H2AX increases in CML and nCML after sequential treatment IM/no drug → GO.....	149
Figure 5-22. Differences on the percentage of cells in each cell cycle phase compared with the NDC after sequential treatment IM/no drug → GO.....	150
Figure 5-23. Treatment with GO before IM reduces cell number in CML and nCML.	154
Figure 5-24. Treatment with IM prevented CML cells from recovering after GO treatment.	155
Figure 5-25. Colony counts after sequential treatment GO → IM/no drug.....	156
Figure 5-26. CFC concentration after sequential treatment GO → IM/no drug.....	157
Figure 5-27. Colony counts after sequential treatment GO → IM/no drug and 6 weeks on culture with stroma cells.	158
Figure 5-28. CFC concentration after sequential treatment GO → IM/no drug and 6 weeks on culture with stroma cells..	159
Figure 5-29. IM treatment after GO reduced the concentration of GO required for increasing the levels of γ H2AX increases in CML.....	160
Figure 5-30. Differences on the percentage of cells in each cell cycle phase compared with the NDC after sequential treatment GO → IM/no drug.....	161
Figure 5-31. Initial exploration of the RNAseq data did not reveal unexpected sources of variability.	163
Figure 5-32. Volcano plots of the single treatment.....	165

Figure 5-33. Volcano plots showing the gene expression changes after the combination treatment.....	166
Figure 5-34. Gene expression changes after the different treatments on the TKIi genes present in the analysis.	167
Figure 5-35. Overlaps between the different treatments and the TKIi genes	168
Figure 6-1. Proposed integration of the TKIi signature with the current literature.	183

Acknowledgements

I would like to thank a number of people who have made the completion of this thesis possible. First, I would like to thank Tessa Holyoake, who was my primary supervisor during the first three years of my project and that, very sadly, is not here to see the completion of the project that she conceived. Thank you for believing in me even in difficult moments and for making my arrival to a new country easier than it would have been without you.

I would also like to thank the rest of my supervisors: Heather, Lisa and Simon. You have done a good job at keeping me going despite the difficult circumstances that we all have lived in the last years. You have been pushing me until the end to make sure all the effort ends in a finished and written thesis. I have not been the easiest student but your perseverance has helped me grow both as a professional and as a scientist. I would specially like to thank Heather for always listening to me when I jumped into the office, at 5:00pm, with some last minute questions about the experiments.

I would also like to thank everyone else at the Paul O’Gorman for the constant support and help. Thank you Alan for your teachings into Scottish and Russian language and culture, for your amazing job keeping the biobank, the technical support and for keeping me in mind every time a new sample arrived. Thank you Jennifer for helping with the SFM, and for resolving all my questions regarding FACS and colour panels. And thank you Joana for your very fast and efficient job at annealing the reads of my RNAseq experiment.

However, I would like to say a special thank you to Russell, Chinmay, Caroline and Lorna. You have been a constant help in all those moment when I thought that I reached the limit. You have helped me to keep going, look at the project with a brighter perspective and see that it is not an impossible task. Russell, you have been a constant help every time I had a problem with my code. I will definitely miss our before-home (always late) chats. Chinmay: thank you for your selfless help in the lab. You made possible for me to go on holidays more than one time and you have been a very good friend. Lorna: we have followed a very similar path during these years. It has been great to have a journey companion like you. It has definitely made the completion of this PhD easier.

I would also like to thank all the group leaders from the Paul O’Gorman for the constant input and very useful suggestions on my project. Special thanks to Mhairi for dealing with all the paper work with Pfizer and to David for your ideas and the relaxed late drinks in Estoril.

I would like to show all my gratitude to Bloodwise for funding my project and giving me the opportunity of doing this exciting research as well as for giving me the opportunity to present my work every year in London.

Thank you to all my friends. To Máté for all the good intentions of cycling around and going to the gym but also for helping me to socialise when I only had work. To my friends from Zizur, especially to Miren for giving me a home every time I visit Edinburgh. And thank you to all my friends from BQ rules, who make every day easier even from the distance and keep me closer to home.

No me olvido de mis padres. Gracias por enseñarme a salir adelante en los momentos difíciles, y a luchar por lo que quiero. No podría haber llegado hasta aquí sin el sacrificio económico durante los años de la carrera y el máster. Siempre habéis creído en mí y a pesar de que estamos lejos, sé que compartís cada logro y cada fracaso. Mamá, eres un ejemplo de que trabajando duro se puede salir de los mayores problemas. Estoy muy orgulloso de ti. Papá, tengo que decir que tú también has tenido pasar pruebas muy duras y de ti he aprendido a no echarme atrás por nada. Aunque pueda sonar paradójico me has enseñado a ser un hombre y creo que no podría haber tenido mejor ejemplo. Mikel, eres un PP con malas pulgas pero no te cambiaría por nadie. Haces que ir de vacaciones a un pueblo con peor clima que Glasgow sea un lujo. Gracias por querer a mi madre, por apoyarnos en los momentos difíciles y por servirnos esos tomates tan ricos todos los años. Gracias por todo. Os quiero mucho y sin vosotros no habría llegado hasta donde estoy.

Gracias abuelito por quererme tanto mientras estuvimos juntos. Tu partida fue el desencadenante de mi decisión de dedicarme a la investigación contra el cáncer, así que tú eres el responsable de que haya llegado hasta Glasgow.

And last, I cannot thank you enough, Arina, моя девушка. Thank you for being patient every time (always) that I left the lab late. Sorry for all those rushing times before going on holiday. I could not have done all of this without you. You have always believed in me and you have been an example of hard-work and success. Te quiero. я тебя люблю.

Related publications

1. JACKSON, L., **GÓMEZ-CASTAÑEDA, E.**, JØRGENSEN, H. G., HOPCROFT, L. E. M., ROGERS, S., HOLYOAKE, T. L. & HUANG, X. 2017. Oncogene Addiction in Chronic Myeloid Leukaemia. *eLS*. John Wiley & Sons, Ltd.
2. HAYATIGOLKHATMI, K., PADRONI, G., SU, W., FANG, L., **GÓMEZ-CASTAÑEDA, E.**, HSIEH, Y. C., JACKSON, L., HOLYOAKE, T. L., PELLICANO, F., BURLEY, G. A. & JØRGENSEN, H. G. 2017. Investigation of a minor groove-binding polyamide targeted to E2F1 transcription factor in chronic myeloid leukaemia (CML) cells. *Blood Cells Mol Dis*.
3. O'CONNOR, C., YALLA, K., SALOMÉ, M., MOKA, H. A., **GÓMEZ-CASTAÑEDA, E.** & KEESHAN, K. 2018. Trib2 expression in granulocyte-monocyte progenitors drives a highly drug resistant acute myeloid leukaemia linked to elevated Bcl2. *Oncotarget*.

Publications under preparation:

1. KINSTRIE, R., HORNE, G., MORRISON, H., IRVINE, D., MUNJE, C., **GÓMEZ-CASTAÑEDA, E.**, DUNN, D., CASSELS, J., SCOTT, M., HOLYOAKE, T., WHEADON, H. & COPLAND, M. CD93 is selectively upregulated on CML stem cells and identifies a quiescent population which persists after tyrosine kinase inhibitor therapy. In preparation.
2. **GÓMEZ-CASTAÑEDA, E.**, MUNJE, C.R., BITTENCOURT-SILVESTRE, J., HYDE, R., COPLAND, M., ROGERS, S., HOPCROFT, L.E.M. & JØRGENSEN, H.G. The discovery of a BCR-ABL1 TK independent signature in CML LSCs opens new avenues for the treatment of CML and predicts response to TKI treatment. In preparation.

Author's declaration

I declare that all the data presented in this thesis is my own work, unless otherwise stated and has not been submitted in any form for the award of any other degree.

Eduardo Gómez Castañeda

September 2018

List of abbreviations

aHSCT	Allogenic stem cell transplantation
AIAY	Aggressive-Indolent Agnes Yong (dataset)
AP	Accelerated phase
BC	Blast crisis
BOS	bosutinib
CCyR	Complete cytogenetic response (<0.5% t(9;22) positive metaphases) (Baccarani et al., 2009)
CFC	Colony forming cell
CML	Chronic myeloid leukaemia
CMLDV	Chronic myeloid leukaemia David Vetrie (dataset)
CMLMC	Chronic myeloid leukaemia Mhairi Copland (dataset)
CP	Chronic phase
DAS	dasatinib
E	Erythrocyte
G	Granulocyte
G-CSF	Granulocyte colony stimulating factor
GEMM	Granulocyte, erythrocyte, macrophage and megakaryocyte
GM	Granulocyte macrophage
GO	Gemtuzumab-ozogamicin
GO→IM	Sequential treatment of 72h of GO followed by 72h of IM
HSC	Haematopoietic stem cells
IM	imatinib
IM→GO	Sequential treatment of 72h of IM followed by 72h of GO
LSC	Leukaemia stem cell
LTC-IC	Long term culture initiating cell
M	Macrophage
MFI	Median fluorescence intensity
MMR	Major molecular response ($\leq 0.1\%$ of BCR-ABL1 transcripts compared with ABL1) (Baccarani et al., 2013)
nCML	Non-chronic myeloid leukaemia
NDC	No drug control
NIL	nilotinib
PC	Principal component
PCA	Principal component analysis
PON	ponatinib
RNRMW	Responders no-responders McWeeney (dataset)
TKI	Tyrosine kinase inhibitors
TKIFP	Tyrosine kinase inhibitors Francesca Pellicano (dataset)
TKIi	Tyrosine kinase inhibitor independent
ρ	Spearman's rho

1 Introduction

1.1 Haematopoiesis

The blood is a fluid and dynamic tissue that has an important role in the homeostasis of the different organs and tissues of the body, maintaining constant levels of oxygen and pH as well as other salts, and neutralizing foreign elements. The mature cells of the blood have a short life and over 3×10^{11} new blood cells need to be generated in an adult human every day (Notta et al., 2016). This constant repopulation is performed by the haematopoietic stem cells (HSCs), a population of multipotent stem cells that are capable of differentiating into any haematopoietic cell type and have self-renewal capacity (Akashi et al., 2000, Notta et al., 2016).

Studies on the effect of radioactivity in mice led to the finding that transplanting bone marrow from a non-irradiated healthy mouse into a lethally irradiated mouse was able to rescue the mouse and reinitiate haematopoiesis, and that this protective effect was performed by the cellular component of the bone marrow (Ford et al., 1956). Further engraftment studies in lethally irradiated mice found that some cells, the colony forming cells (CFC), form colonies in the mouse spleen when transplanted and that these colonies contain committed mature blood cells of the myeloid lineage (Till and McCulloch, 1961). Additionally, analysing unique genome abnormalities induced by radiation it was possible to confirm the clonal nature of those colonies (Becker et al., 1963). Study of these colonies revealed that additional CFCs were contained in each colony, suggesting that CFCs have self-renewal capacity (Siminovitch et al., 1963). Additional work investigating the presence of repopulating cells in the thymus and the lymph nodes revealed that those cells contained the same chromosomal abnormalities as the CFCs found in the spleen, suggesting that lymphoid cells derive from CFCs or that both types of cells share a common progenitor (Wu et al., 1968). These experiments led to the idea of the existence of a type of cell that is able to generate a progeny that belongs to all the myeloid and lymphoid cell types and that can replicate itself (self-renewal): that is the HSC. However, it was not until 1988 that this cell population was isolated in the mouse by fluorescence-activated cell sorting (FACS), identified by the cell surface markers $\text{Thy-1}^{10}\text{Lin}^{-}\text{Sca-1}^{+}$ (Spangrude et al., 1988). Transplanting just 30 of these HSCs was able to rescue lethally irradiated mice and generated daughter cells belonging to all myeloid and lymphoid cell types (Spangrude et al., 1988).

The functional characterisation of human HSCs required the development of *in vitro* assays that could re-create the experimental conditions observed in mouse transplant

experiments in such a way that both the ability to generate colonies and self-renewal were measured. First, an *in vitro* system in which the cells were cultured on methylcellulose for two weeks allowed CFC to form colonies of the myeloid lineage (Hara and Ogawa, 1978). As HSCs are contained within the CFC, it was assumed that HSCs would also have the capability of forming colonies on methylcellulose media. However, for characterising the HSCs in this way it was necessary to purify them first. It was observed that most CFCs die after approximately 4 weeks of *in vitro* culture supported by bone marrow stroma yet a non-dividing population persists (Andrews et al., 1986, Eaves et al., 1986). Harvesting the remaining bone marrow cells (discarding the stroma cells) and placing them on methylcellulose to analyse the colony forming potential of the remaining cells allowed quantification of the number of long term culture initiating cells (LTC-IC). These cells were associated with the HSCs because of their long-term self-renewal/survival capability and their capacity to generate colonies (Sutherland et al., 1989). A couple of years later, it was revealed that human HSCs are contained within the $CD34^+CD38^-$ population (Terstappen et al., 1991).

Although mouse $Thy-1^{10}Lin^-Sca-1^+$ were highly enriched for HSCs, the population was still heterogeneous and only 25% of the cells could reconstitute the bone marrow long-term (Morrison and Weissman, 1994). Further study in this heterogeneity of HSCs revealed that there are distinct functional groups of HSCs that can be identified by their long-term engraftment capacity (long-term HSCs; LT-HSCs) and others with shorter time engraftment (still reconstituting all the haematopoietic lineages) but higher proliferation, called short-term HSCs (ST-HSC) and multipotent progenitors (MPP) (Morrison and Weissman, 1994). The finding that HSCs differentiate into committed univariate progenitors (Ogawa, 1993) and the latter discovery of cells that can differentiate to all the cells of either the lymphoid lineage, i.e. common lymphoid progenitors (CLPs), (Kondo et al., 1997) or myeloid lineages, i.e. common myeloid progenitors (CMPs), (Akashi et al., 2000) suggested that haematopoiesis happens in a hierarchical manner. Thus, this model suggested that haematopoiesis is a hierarchical process where the cells lose self-renewal and plasticity with each fate decision and that each intermediate state or population is homogeneous both in function and in the expression of surface markers (Orkin, 2000). Hence, the primitive HSCs would lose self-renewal capacity when becoming MPPs which then would differentiate into the oligopotent progenitors CLP or CMP. CLPs would then differentiate into unipotent progenitors of the lymphoid lineage (T-cells, B-cells and natural killers) and the CMPs would differentiate into megakaryocyte-erythrocyte progenitors (MEP), which would differentiate into platelets and erythrocytes, or GMPs,

which would differentiate into granulocytes, monocytes and dendritic cells (Ogawa, 1993). The model is summarised in Figure 1-1A.

Despite the hierarchical model working well for marker-homogenous populations, it is not the case when investigating single-cell populations (Notta et al., 2016, Guo et al., 2013). One of the first publications to publicly challenge the model described a lymphoid primed multipotent progenitor (LMPP) that was able to generate as progeny all the cells from the lymphoid lineage in addition to both granulocytes and monocytes but did not contain platelets/megakaryocytes or erythrocytes (Adolfsson et al., 2005). This did not fit with the previous model as it was contrary to the concept of CMP, a myeloid progenitor for granulocytes, monocytes, platelets and erythrocytes that cannot generate lymphoid offspring (Akashi et al., 2000). Subsequent publications, have maintained the same direction. Gene expression analysis of single cells have shown that GMPs have a gene expression profile that is closer to CLP than MEP, supporting the idea that MEP branches out of the model before the LMPP (Pronk et al., 2007, Guo et al., 2013). Additionally, LT-HSCs seem to be primed for platelet differentiation as LT-HSCs express *VWF* but not a single LT-HSC has been found to be primed for any other cell fate (Sanjuan-Pla et al., 2013, Carrelha et al., 2018, Notta et al., 2016). Analysis of the capacity of single cells to form colonies with more than one cell type revealed that most cells classified as oligopotent (that can generate an offspring containing cells of more than one type but are already committed, such as GMPs) were only able to generate cells of one type (Notta et al., 2016). This suggested that in adult haematopoiesis there are only two stem/progenitor stages: HSCs and unipotent progenitors (those whose offspring is composed of only one cell type) (Notta et al., 2016). A later report that combined single-cell functional assays, RNAseq and cell surface markers performed in haematopoietic cells of 2 individuals found that there is (I) a *cloud* that contains the multipotent cells, that is, the HSCs, MPPs and LMPP, that retain multipotency and the ability to differentiate into any haematopoietic cell and (II) unipotent progenitors (and a few rare dipotent progenitors) which are already committed towards a cell type (Velten et al., 2017). The group of multipotent cells was referred as a *cloud* because of the high degree of gene expression similarity between the different multipotent cells despite their lineage priming, in opposition to the clear distinction between the different unipotent progenitors (Velten et al., 2017). Those unipotent progenitors however are not an isolated cell type but also form a continuum that accumulate changes from the multipotent cloud until the final differentiated progeny (Velten et al., 2017, Tusi et al., 2018).

The idea of a long-living group of progenitor cells is also supported by undisturbed haematopoiesis studies. The study of adult haematopoiesis in transposon tagged mice allowed identification of the cellular origin of the blood cells in each mouse as each cell would have a unique label (Sun et al., 2014). This system provided a model where no cellular stress was induced and there was no need for the cells to invade other tissues (engraft), a process that can introduce bias towards a certain type of cells in transplantation studies (Sun et al., 2014). Analysing global haematopoietic progeny, the contribution of LT-HSCs to blood production seems to be very limited and instead, multipotent progenitors (MPP) and ST-HSCs, which are long-living, are driving life-long haematopoiesis after birth (Sun et al., 2014).

These recent publications have rebuilt the haematopoietic differentiation hierarchy into a continuum where no unique states are present (Tusi et al., 2018, Velten et al., 2017, Notta et al., 2016, Carrelha et al., 2018). In this model, platelets derive from the most primitive HSCs (Carrelha et al., 2018) and then another two branches, one leading to basophils, eosinophils, mast cells and erythrocytes, and another leading towards myeloid (neutrophils, monocytes and dendritic cells) and lymphoid differentiation (Notta et al., 2016, Tusi et al., 2018, Velten et al., 2017). The model is summarised Figure 1-1B.

Originally it was believed that haematopoietic differentiation was directed by cytokines (Ogawa, 1993). However, the non-specific pathways activated by the cytokine-receptor signalling cascade led this cytokine-driven differentiation model to be questioned (Socolovsky et al., 1998). In fact, it was shown that lineage-associated receptors are not necessary for cell differentiation into that particular cell type but to promote progenitor survival (Socolovsky et al., 1998). One example of this was the expression of the chimeric prolactin receptor that contained the pro-survival cytoplasmic domain of the erythropoietin receptor (EPOR) in erythroid progenitors that had a non-functional EPOR. These cells were able to proliferate and differentiate after signal transduction induced by prolactin but not when the pro-survival cytoplasmic domain of the chimeric receptor was non-functional (Socolovsky et al., 1997). This led to the need to find a control mechanism in haematopoietic cells to control cell fate and differentiation. Finding that the absence of some transcription factors impaired haematopoiesis suggested that it is the presence of certain transcription factors that is actually regulating the process (Socolovsky et al., 1997, Orkin, 2000).

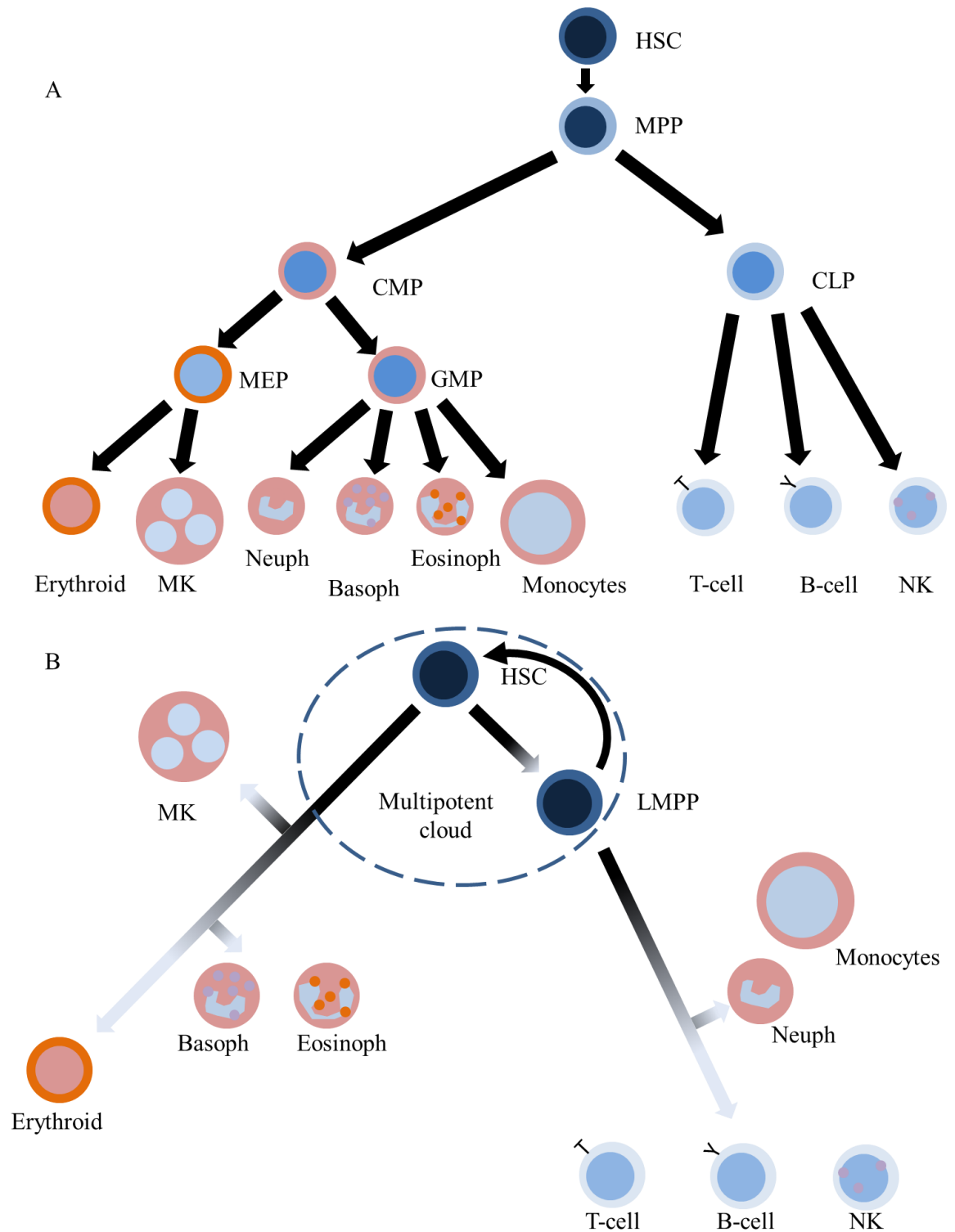


Figure 1-1. Different haematopoiesis models. Gradient of blue represents multipotent capacity, with darker blue having more potency (capacity to differentiate into different cell lineages) and fainter blue meaning a more committed phenotype. Red colour represents the cells that have traditionally been identified with the myeloid lineage. (A) Summarises the traditional hierarchical model with discrete cell populations that have a homogeneous phenotype and expression of cell surface markers. (B) The new model identified a continuum of HSCs and MPPs that retain multipotency and that during commitment they suffer gradual changes (represented by the colour gradient in the arrows) instead of differentiating into discrete cell populations.

It was believed that different transcription factors are expressed in HSCs and that their presence maintains an open conformation of the chromatin of those genes that are needed for cell differentiation (Orkin, 2000). Although their presence is needed for both, maintaining the self-renewal HSC phenotype or to differentiate into the different haematopoietic cell lineages, it is the concentration or proportion of each of them that would determine the fate of each individual cell (Orkin, 2000, Orkin and Zon, 2008). Many of these transcription factors were discovered to have a role in normal haematopoiesis after being found as part of fusion proteins in different types of leukaemia. One such transcription factor is PAX5. Mice deficient in PAX5 were able to generate pro-B cells, a precursor of B-cells, but these cells were unable to differentiate into mature B-cells (Nutt et al., 1999). Instead, PAX5 deficient pro-B cells expressed genes from different haematopoietic lineages and differentiated into functional macrophages, osteoclasts, dendritic cells, granulocytes and natural killer cells (Nutt et al., 1999), which is in line with the existence of a LMPP progenitor (Adolfsson et al., 2005, Velten et al., 2017, Tusi et al., 2018). However, upon reconstitution of *Pax5* expression, the cells silenced the non B-cell genes and differentiated into B-cells (Nutt et al., 1999). This showed that PAX5 is needed for negative regulation of other cell fates. GATA1's effect was shown to be dose dependent as myeloid cells expressing modest levels of GATA1 protein differentiated into eosinophils while cells with high levels of GATA1 differentiated into platelets and erythrocytes (Kulesa et al., 1995). GATA1 was also shown to mutually antagonise PU.1 (Rekhtman et al., 1999, Zhang et al., 2000), another dose dependent transcription factor that favours B-cell differentiation at low concentration and macrophage differentiation at high concentrations (DeKoter and Singh, 2000). Both GATA1 and PU.1 proteins interact preventing each other to bind to their DNA target regions and, therefore, inhibiting each other's activity. The association of the amino terminus of PU.1 with the carboxy-finger of GATA1 blocks GATA1's function, probably by interfering with its DNA recognition site (Zhang et al., 2000). GATA1's inhibition of PU.1 seems to be through the interaction of the GATA1 carboxy finger with the ETS domain of PU.1, which is PU.1 DNA-binding domain (Nerlov et al., 2000).

However, transcription factors are not only necessary for guiding cell differentiation but also for its induction or the maintenance of a stem cell phenotype. Increased proliferation can lead to stem cell exhaustion, when the bone marrow is unable to generate the necessary number of blood cells (Cheng et al., 2000, Bernitz et al., 2016). Therefore, controlling stem cell proliferation and fate is crucial for viability of the whole organism. For example, the IK6 dominant negative isoform of IKAROS, a key factor in lymphoid (Nakayama et al.,

2000) and erythroid differentiation (Dijon et al., 2008), has been found to increase the proportion of CD34⁺CD38⁻ human cells transplanted in mice while reducing the amount of erythroid cells and increasing the number of B and myeloid cells (Beer et al., 2014). These IK6⁺ HSCs did also present a higher engraftment potential and did engraft secondary recipients (Beer et al., 2014). This would suggest that IKAROS is a negative regulator of stem cell maintenance. Loss of function (LOF) mutations of *RUNX1* in HSCs have been observed to increase the number of HSCs and the cells of the erythroid and myeloid lineages while negatively affecting the development of platelets and lymphoid cells (Gowney et al., 2005). In HSCs, LOF of *RUNX1* decreases cell metabolism, ribosomal activity and proliferation while increasing cell resistance to stress and DNA damage (Cai et al., 2015). This increased survival in cells with *RUNX1*^{LOF} and provided a selective advantage over normal HSCs. MYC is also an important transcription factor involved in the regulation of HSCs differentiation. Reduced activity of MYC has been associated with impaired ability of HSCs to differentiate, which increase the expression of cell adhesion markers, such as N-cadherin (Wilson et al., 2004), and accumulate in their niche (Laurenti et al., 2008).

Although most transcription factors in haematopoiesis seem to participate in the positive regulation of cell differentiation, it is likely that stem cell maintenance and self-renewal are also positively regulated by a set of transcription factors. The effects of the transcription factor ERG in HSCs have been studied in the last years founding a strong relation between the levels of ERG expression and HSC activity (Loughran et al., 2008, Ng et al., 2011, Taoudi et al., 2011, Xie et al., 2017, Knudsen et al., 2015). Mice carrying 2 copies of an *Erg* mutant with no transactivation capacity (*Erg*^{Mld2}) die *in utero* as they fail to maintain definitive haematopoiesis (Loughran et al., 2008, Taoudi et al., 2011). However, observation of the haematopoietic system during development showed that ERG is not necessary for HSC production but to their maintenance (Taoudi et al., 2011). *Erg*^{Mld2/+} mice present lower levels of platelets and leukocytes in parallel with a slightly reduced HSC compartment (Loughran et al., 2008, Xie et al., 2017). Although steady haematopoiesis was not severely affected, *Erg*^{Mld2/+} HSCs had poorer engraftment capacity and were outcompeted by *Erg*^{+/+} even when they were transplanted in much higher proportion (Loughran et al., 2008). Furthermore, *Erg*^{Mld2/+} HSCs were impaired during stress-induced haematopoiesis and lost self-renewal capacity, leading to a drain of the HSC pool (Ng et al., 2011). HSC exhaustion was also found in transplantation studies in a double inducible knock out mouse model (Knudsen et al., 2015). Interestingly, it was found that ERG is an upstream regulator of other haematopoietic transcription factors,

including *Gata2* and *Runx1*, which shows the role of ERG as a HSC master regulator (Taoudi et al., 2011). Additionally, ERG has been found to bind to MYC motifs and directly control a set of MYC regulated genes, although it does not regulate *MYC* expression (Knudsen et al., 2015). This repressive role in MYC-regulated genes expression explains the lack of self-renewal capacity and increased differentiation in ERG deficient cells as reduced ERG activity leads to increased levels of MYC-regulated genes (Knudsen et al., 2015). This activates a high MYC activity like phenotype, which induces proliferation and cell differentiation (Wilson et al., 2004). Interestingly, HSC phenotype can be partially rescued using bromodomain and extra terminal protein inhibitors (BETi), compounds that reduce MYC activity (Knudsen et al., 2015). Additionally, induction of *ERG* gene expression in combination with induction of gene expression of *HOXA5*, *HOXA9*, *HOXA10*, *LCOR*, *RUNX1* and *SPI1* has been found to convert haemogenic endothelium into HSCs with engraftment and both myeloid and lymphoid repopulating capacity (Sugimura et al., 2017). Interestingly, ERG has been found to compete with unphosphorylated STAT5 for the accession to megakaryocytic genes' promoters in HSCs in the absence of cytokine signalling (and therefore, absence of phosphorylation of STAT5 by JAK2) (Park et al., 2016). The affinity of ERG for megakaryocytic promoters may be related with the platelet/megakaryocytic-priming of primitive stem cells previously discussed (Sanjuan-Pla et al., 2013, Notta et al., 2016, Carrelha et al., 2018), which further supports the role of ERG as a positive stem cell regulator.

1.2 Cancer

Cancer is a group of diseases characterized by abnormal gene expression (Croce, 2008). Each tumour has its own characteristics that depend on the tissue of origin, the cell type affected, the alterations that are carried by the cells and the patient. However, tumours conserve some common qualities or capacities. These capacities are also known as the hallmarks of cancer and confer to the tumour adaptive advantages (Hanahan and Weinberg, 2011). The main capabilities are resistance to cell death, independence of growth factors, resistance to growth suppression, induction of angiogenesis, replicative immortality, invasiveness and metastasis capacity, promotion of inflammation, resistance to immune destruction, alterations in metabolism and genomic instability. The origin of these capabilities is usually found in mutations or chromosomal abnormalities (that may cause a wide alteration of the transcriptome), but can also be produced because of environmental conditions in the absence of mutations (both microenvironment and non-microenvironment). Mutations can be inherited mutations, such as *BRAC1* in breast cancer or *MSH* genes in colon and endometrial cancer (germline mutations), or acquired later in life due to failure in the cell replication machinery or the contact with carcinogenic elements, such as radiation or tobacco.

Cancer development and its cellular organization, including hierarchy, have been key questions in biomedical research for several years and they remain poorly understood. Although originally it was believed that tumours were homogenous groups of cells this idea is no longer accepted in the scientific community, replaced by other hypotheses that base the cancer development and persistence in tumour heterogeneity. This heterogeneity refers to non-cancerous cells in the tumour, the microenvironment, but also to the hierarchy of the cancerous cells. Two main models have been proposed: hierarchical and stochastic (Wang et al., 2014). The hierarchical model defends the existence of a pool of cancer stem cells (CSC) that act similarly to normal stem cells, repopulating the tissue (the tumour in this case) and presenting a set of survival and anti-apoptotic signals that allow them to evade treatment related death. According to this model the original cell affected by the cancerous alteration (e.g. chromosomal aberration, mutation, and epigenetic deregulation) is a tissue stem cell and CSC would be the only cells with the capacity to repopulate the tumour or to metastasize, so targeting CSC and killing all of them would cure the disease. On the other hand, the stochastic model proposes that cancerous cells are plastic and are able to alter their fate, switching between CSC status to a more differentiated one and *vice versa*. According to this model every cell in a tumour would have the capacity to repopulate the tumour and to metastasize, inferring the requirement to

target both CSC and non-stem cancer cells (NSCC) when treating in order to cure the disease.

Recent studies (Huntly et al., 2004, Wu et al., 2007, Wang et al., 2014), despite explaining how the tumour develops in a certain type of cancer, are not able to generate a unified model that would help us to understand cancer as a whole. It has been found that different oncogenes have different influences in cell fate regulation, proliferation and self-renewal (Wu et al., 2007, Wang et al., 2014). These results correlate with other previous well-accepted concept of cancer requiring two different types of mutations, one that blocks differentiation and another that promotes cell survival and proliferation. The requirement for these two types of mutations can be easily observed in acute leukaemias (Passegue et al., 2003). Usually cells first acquire the mutation that confers survival and proliferation advantage as a block in differentiation without survival signals would lead to apoptosis. These two steps are also known as initiating and promoting mutations (Croce, 2008).

Haematological malignancies or leukaemias are a group of clonal diseases that affect the cells of the different haematopoietic lineages. These diseases are classified as lymphoid or myeloid depending on the affected lineage and as chronic or acute according to their aggressiveness. Chronic leukaemias are characterized by the increase in the production of a mature cell type in the bone marrow while in the acute leukaemias the proliferating cells are arrested in an immature stage (Hoffman et al., 2012). In acute leukaemias, the patient suffers a rapid increase in the number of immature cells or blasts, in a similar way to the final stages of chronic leukaemias. These cells have morphological abnormalities and are unable to differentiate and are able to interfere with normal haematopoiesis and to infiltrate in other organs (Estey, 2013).

1.2.1 Chronic Myeloid Leukaemia (CML)

1.2.1.1 Clinical presentation of CML

CML is a clonal disease that affects the myeloid lineage (Apperley, 2015) and is characterized by the presence of the reciprocal translocation t(9q;22q) (Rowley, 1973) which results in the presence of the fusion gene *BCR-ABL1* (Shtivelman et al., 1985). CML presents in three different phases: chronic phase (CP), accelerated phase (AP) and blast crisis (BC) (Baccarani et al., 2013) with most patients being diagnosed during CP. At diagnosis in CP, patients usually present an accumulation of granulocytes and progenitors in the bone marrow and in the blood (Shtivelman et al., 1985) accompanied with malaise and splenomegaly (Apperley, 2015). The disease progresses to AP and BC, achieving an acute myeloid leukaemia (AML)-like state, with increasing amounts of immature cells,

both in the blood and in the bone marrow. These cells can have either myeloid or lymphoid phenotype (Apperley, 2015).

CML is a rare disease predominantly found in older age groups (the median age of diagnosis is 57-60 years) (Sokal et al., 1984). However, paediatric cases have also been recorded (Hijiya et al., 2016). It is established that CML accounts for 20% of all leukaemias in adults (UICC, 2014) and 2% of all leukaemias observed in children under 15 years old (Ries et al., 1999). European CML registries suggest that annual CML incidence ranges between 0.7 and 1.0 for every 100,000 people (Hoglund et al., 2013) while some US-based cohorts report the CML incidence varying between 1.4 and 2.0 (Chen et al., 2013).

A series of potential risk factors have been identified as in CML. Probably, exposure to ionising radiation and increasing age are the two factors with higher impact in the risk of development of CML. Hiroshima atomic bomb survivors were reported to have a higher incidence of CML the closer they were to the epicentre of the explosion (Heyssel et al., 1960). However, this is an extreme example of radiation exposure and the general population is not at risk of radiation exposure of that magnitude. Sex may also be regarded as a risk factor in CML as observational studies reveal that CML is found between 20% and 70% more commonly in men than in women (Hoglund et al., 2015, Radkiewicz et al., 2017). This observation was also confirmed by investigation of Japanese atomic bomb survivors. A possible explanation of sex differences in CML incidence may be due to a higher number of target cells at risk in the male body to develop CML and hence an increased risk of oncogenic mutations (Radivoyevitch et al., 2014). Moreover, male sex association with a poorer prognosis may be partially explained by a historically higher alcohol and tobacco consumption in men (Radkiewicz et al., 2017). Occupational exposure to benzene has also been associated with CML development (Vlaanderen et al., 2012, Adegoke et al., 2003). However, the studies are based on small number of people therefore the evidence is sparse. Smoking may also contribute to the arising of CML (Musselman et al., 2013) as tobacco smoke contains both carcinogenic chemicals and benzene. However, no enough information is available to confirm this hypothesis. Nonetheless, smokers tend to have poorer survival rates than non-smokers (Lauseker et al., 2017).

1.2.1.2 Molecular and cellular characteristics of CML

The pathognomonic characteristic of CML is the presence of the BCR-ABL1, a protein which main isoform is 210KDa, although isoforms of 190KDa and a 230KDa have also been described (Deininger et al., 2000). The pathologic effect of this protein is due to the

deregulation of the ABL1 tyrosine kinase activity, which is constitutively activated (Zhao et al., 2002). Constitutive activation of the tyrosine kinase domain of ABL1 seems to be mediated by the oligomerization of the fusion protein (Zhao et al., 2002). The N-terminal region of the BCR segment contains an oligomerization domain that allows the dimerization of homodimers of BCR-ABL1 (tetramers) forming a very hydrophobic core. The proximity of the tyrosine kinase domains allows tyrosine auto-phosphorylation and, therefore, activation of the kinase domain (Deininger et al., 2000). The oligomerisation also promotes protein-protein interactions with other proteins containing SH2 domains (Deininger et al., 2000). This alters the signalling pathways in the affected cells through an increased phosphorylation led by the deregulated TK activity and by the increased docking of proteins containing SH2 domains. In fact, BCR-ABL1 is implicated in the signal transduction in many proliferative and survival pathways, including PI3K/AKT (Skorski et al., 1995) and Ras/Rho (Pendergast et al., 1993), JAK/STAT (Carlesso et al., 1996, Xie et al., 2001) and MYC pathways (Xie et al., 2002). The signalling alterations induced by BCR-ABL1 have been found necessary and sufficient for the development of CML. Using *in vivo* mouse models it was demonstrated that the expression of *BCR-ABL1* in the HSC population promotes a CML-like disease in these mice (Koschmieder et al., 2005). However, BCR-ABL1 has been found in circulating blood cells of healthy individuals at very low levels (Biernaux et al., 1995), suggesting that not every cell carrying the mutation initiates the disease.

It has actually been described that BCR-ABL1, in contrast with other oncogenes, is not able to activate a self-renewal program in transfected cells (Huntly et al., 2004). In fact, when BCR-ABL1 was expressed in a committed population the leukaemic clone exhausted (Huntly et al., 2004). The description of a CML cell population that resembles both the cell surface markers and the biological properties of HSCs, including quiescence, cell renewal and colony forming capacity (Holyoake et al., 1999) suggests that BCR-ABL1 has to originate in a HSC for maintaining the disease. This idea is further supported by the previous finding that all myeloid lineage cells (erythrocytes, granulocytes, platelets and monocytes) can be generated by the leukaemic clone in CML (Fialkow et al., 1967, Fialkow et al., 1977). Furthermore, another publication reported that a CML-like disease could be observed running in the background in patients suffering from BCR-ABL1⁺ acute lymphoid leukaemia (ALL), suggesting that both the ALL and the CML cells originate from a common progenitor able to generate cells of different lineages (Hovorkova et al., 2017).

1.2.1.3 Treatment of CML

CML treatment has experienced important changes in the last decades, going from the use of alkylating agents to the use of the revolutionary tyrosine kinase inhibitors (TKI). In the early 1980s, the standard treatment for CML was the use of busulfan, a DNA alkylating agent, and hydroxyurea, a myelosuppressive agent (Silver et al., 1999). The use of these compounds only achieved modest success and 45% patients treated with them achieved 5 years survival after the diagnosis (Hehlmann et al., 1994). Moreover, the adverse effects of the drugs were frequent and severe. For instance, busulfan was found to induce hepatic, pulmonary and cardiac fibrosis and hydroxyurea was associated with normal blood myelosuppression and reversible renal and liver dysfunctions (Rushing et al., 1982, Tsukagoshi, 1992).

Interferon- α (IFN α), an important immunogenic cytokine, was first introduced as a treatment option for CML in 1983 (Talpaz et al., 1983, Gutterman, 1994). It was found that IFN α was able to promote more durable cytogenetic response than previously used anticancer drugs (Talpaz et al., 1987). IFN α treatment increased the rate of patients achieving 5-year survival to 59%, an increase of around 15% compared with previous treatments based on busulfan and hydroxyurea (Ohnishi et al., 1995, Hehlmann et al., 1994, Chronic Myeloid Leukemia Trialists' Collaborative, 1997). However, IFN α treatment presents disturbing side effects, especially during long term treatment, including flu-like syndrome and even depression and combination treatments with other drugs were tested in order to reduce the dose of IFN α . Therefore, the combination of IFN α with hydroxyurea (Kantarjian et al., 1993) and cytarabine (Arthur and Ma, 1993, Kantarjian et al., 1999) allowed reduction of the dose of IFN α , hence reducing treatment induced toxicity.

Allogeneic stem cell transplantation (aHSCT) has been widely used since 1980s as the only truly curative method for CML (Thomas et al., 1986, Baccarani et al., 2013). The transplant offers a cure due to its conditioning regimen and its immunogenic effect against the leukaemic cells through the donor's T-cell response. However, aHSCT is associated with substantial risks of transplantation-related mortality (Sawyers, 1999) and currently, it is only recommended for advanced stages of CML, or for patients with resistance to multiple TKIs (Baccarani et al., 2013). The European Group for Blood and Marrow Transplantation (EBMT) score is also used as a risk assessment for aHSCT (Gratwohl, 2011). The score ranges from 0 to 7 and is based on the age of the patient, stage of the disease, the donor/recipient sex, donor type (sibling/unrelated) and time from diagnosis with 0 being the best score (better post-transplant survival) and 7 being the worst (Gratwohl et al., 2009). Thus, despite aHSCT being the only known curative treatment for

CML, the risk associated with it displaces it to a third-line treatment for patients who progress to AP or BC or those with resistance to more than one TKI (Baccarani et al., 2013).

The current standard of treatment in CML is the use of TKIs, a group of small molecule inhibitors that are able to bind to the catalytic pocket of the TK domain of BCR-ABL1 and prevent its interaction with ATP (Druker et al., 1996, Zabriskie et al., 2014). The first compound to be available for its use in the clinical practice was imatinib (IM), which showed a very specific targeting of the BCR-ABL1 TK activity, although the KIT receptor and PDGFR have also shown inhibition upon treatment with IM (Heinrich et al., 2000). Clinical trials comparing IM to the previous treatment consisting of IFN α and cytarabine showed a much better cytogenetic response on the patients treated with IM and also an increase in the survival rate after 18 months of treatment (O'Brien et al., 2003). However, IM was found not to eradicate the leukaemia stem cells (LSCs) over treatment *in vitro* (Graham et al., 2002) and patients suffering from CML need to be on treatment for the rest of their lives for controlling the disease. Only a small subset of patients presenting sustained major molecular response (MMR), accounting for 4% of the CML patients treated with TKIs, are able to discontinue the treatment without suffering a relapse (Mahon et al., 2010).

The appearance of side effects and TK domain mutations that disturb the binding of IM to the catalytic pocket created a clinical need for treating those patients that do not benefit from IM (Baccarani et al., 2013, Zabriskie et al., 2014). This led to the development of new generations of TKI, which included nilotinib (NIL), dasatinib (DAS), bosutinib (BOS) and ponatinib (PON). These new generation compounds can be used for treating the patients who had serious adverse effects with IM and are effective at binding the catalytic pocket of some mutated BCR-ABL1, although PON is the only compound that has shown activity against the BCR-ABL1^{T315I} (Zabriskie et al., 2014). Although these drugs have a higher affinity for the catalytic pocket of BCR-ABL1 than IM, it has been shown that they are still unlikely to eradicate the LSC population in CML patients (Jorgensen et al., 2007, Copland et al., 2006). Therefore, patients suffering from CML have to be on life-long treatment.

1.2.1.4 Prognosis scores in CML

The lack of a risk-free curative treatment also means that it is important to correctly assess the patients that are at a higher risk of not responding correctly to TKI treatment and therefore, developing an advanced disease over time. This means that higher-risk patients

will be under closer monitoring and different therapeutic options (e.g. second generation TKI) may be considered. Current stratification scores are based on cellular and clinical factors. Historically, the Sokal score (Sokal et al., 1984) was the first score to be implemented for the prognostic assessment of patients suffering from CML. This score was developed based on data of patients treated with cytotoxic agents (busulfan) and the final score took into account the age of the patient at diagnosis, spleen size, platelet count and percentage of blasts. The next prognostic score to be developed was the Euro score (Hasford et al., 1998). This score was generated using data from patients treated with IFN α and found age, spleen size, blast count, platelet count, eosinophil count, and basophil count to be good indicators of prognosis. The EUTOS score, the most recent prognosis score to be published, was developed using data from patients treated with IM in different European countries (Hasford et al., 2011). This score uses the spleen size and at time of diagnosis and the percentage of basophils to generate the score. This score has achieved a high specificity, meaning that patients classified as high-risk will benefit from closer monitoring. However, the EUTOS score only has a modest sensitivity both for CCyR and progression free survival (23% and 16% respectively). This means that a big proportion of the patients that would benefit from closer monitoring are actually classified as low-risk.

The low sensitivity of the score has motivated the investigation of molecular signatures that could predict response to TKI treatment in patients suffering from CML. Whole transcriptome gene expression changes between IM responders and non-responders were investigated using unselected CML cells collected at diagnosis (Crossman et al., 2005). However, no differences were found between the two groups of patients. The same research group then investigated if the use of a more primitive cell population (CD34⁺ cells) gene expression could improve the ability to find differences between the two groups of patients (IM responders and non-responders). Using these primitive cells, it was possible to build a gene expression classifier that was able to discriminate patients from both groups and, therefore, to predict their response to TKI (McWeeney et al., 2010). However, the use of this classifier failed to discriminate between patients that respond well to NIL and those who do not (Patel et al., 2018). This highlighted the need to build a molecular classifier able to predict response to a wider set of TKI treatments and not only IM.

The use of single-cell RNA sequencing has provided a powerful tool for investigating gene expression differences. Using this approach it was possible to investigate *BCR-ABL1*⁺ cells in CML patients (Giustacchini et al., 2017). This revealed that gene expression of these cells is different in patients that responded well to TKI treatment than in those who did not.

In contrast, unsupervised analysis of the whole transcriptome did not revealed differences between the two groups of patients (Giustacchini et al., 2017). Additionally, the same publication reported that patients that evolved into blast crisis already presented a distinct clone with blast crisis like gene expression in the samples collected at diagnosis (Giustacchini et al., 2017). Taken together, single-cell RNA sequencing represents a powerful tool that allows both prediction of response to TKI treatment and the risk of an early blast crisis transformation (Giustacchini et al., 2017). However, this technique is still very expensive and technically challenging. Therefore, the development of standardised protocols and a reduction in the price are needed for its routine implementation in the clinic.

1.2.1.5 CML LSC persistence and evidence of a BCR-ABL1 TK independent signature

As mentioned before (1.2.1.3), TKI treatment despite managing CML is not a curative treatment. Thus, prevalence of CML is increasing and treating patients suffering with CML with TKIs, which are expensive, is proving challenging for public health systems and for patients without health insurance (Abboud et al., 2013, Beinortas et al., 2016, Kurtovic-Kozaric et al., 2016). It was not before the availability of generic analogues of IM (2017 in the UK) that TKI treatment became a norm in many countries that could not afford the cost (Beinortas et al., 2016, Lejniece et al., 2017). However, not every patient responds well to IM and the other TKIs are still expensive. Additionally, the persistence of CML cells in the patients is not only a psychological burden to the patients, but also a constant risk of treatment resistance acquisition or disease progression.

This clinical need prompted the study of CML persistence after TKI treatment. It was soon after the introduction of IM in general practice that researchers in Glasgow found that the LSC, which are quiescent, are not eradicated after IM treatment (Graham et al., 2002). In fact, it was found that a higher proportion of cells showed a quiescent phenotype, suggesting that IM has an antiproliferative effect on this cell population (Graham et al., 2002). Furthermore, recent reports have shown that IM may induce the expression of quiescence and self-renewal genes both in CML and normal cells (Charaf et al., 2016, Zhang et al., 2018). Similar effects on the LSC population have been observed also with the second generation TKIs dasatinib (Copland et al., 2006) and nilotinib (Jorgensen et al., 2007). An increase in the number of quiescent CML LSCs has also been confirmed by a single-cell RNA sequencing experiment after TKI treatment compared with the same patients at diagnosis (Giustacchini et al., 2017). This revealed that the use of TKIs is not

enough for the eradication of the CML clone and that other therapeutic avenues should be explored.

As TKIs have an antiproliferative effect on CML LSCs, they fail to eradicate the disease. Therefore, a potential therapeutic option for the eradication of the CML LSCs could be to promote their entrance into cell cycle, which would make them sensitive to TKI treatment. IFN α has been proposed as a candidate medicine for this approach as six patients initially treated with IFN α showed strong responses to IM when they switched treatment (Essers et al., 2009). Similarly, it was observed that the use of intermittent pulses of granulocyte colony stimulation factor (G-CSF) in CML CD34⁺ cells induced them to enter cell cycle and increased their sensitivity to IM *in vitro* (Jorgensen et al., 2006). With these promising results, G-CSF was tested in patients during a phase II clinical trial. This trial showed that the use of G-CSF is not toxic for the patients (Drummond et al., 2009). However, a higher number of patients lost CCyR or MMR in the study group than in the control group taking the standard 400mg/day of IM, which was believed to be caused by the interruption of IM during the G-CSF pulses (Drummond et al., 2009). Thus, this treatment option was not taken forward, although it is still believed that cell cycle entry is a good strategy for sensitising CML LSCs to TKI treatment.

These results increased the interest in the understanding of the underlying causes of CML LSC persistence in an attempt to find new ways to target them. It was especially revealing that research performed in Glasgow showing simultaneous inhibition of BCR-ABL1 activity by both dasatinib and knock-down failed to eradicate the CML LSCs (Hamilton et al., 2012). Inhibition of BCR-ABL1 TK was confirmed by reduction of CRKL and STAT5 phosphorylation, which confirmed that BCR-ABL1 TK is molecularly targeted by TKIs. Additionally, it was found that although the BCR-ABL1 double inhibition significantly reduced the number of CML cells, these were highly enriched for primitive LTC-IC (Hamilton et al., 2012), showing the high resistance of this population to the current treatments. Interestingly it was found that the activity of IM in eradicating CML progenitor cells (CD34⁺CD38⁺) is achieved not only by the inhibition of BCR-ABL1 TK but the simultaneous inhibition of the KIT (Corbin et al., 2013). It was shown that specific inhibition of the BCR-ABL1 TK was rescued by treatment with stem cell factor (SCF). This suggested that similarly to progenitor cells, LSCs may be rescued from TKI mediated cell death by an unknown mechanism. This was supported by the finding of molecular pathways that are differentially expressed in CML cells that persist TKI treatment, such as the intrinsic β -catenin pathway, which suggests that CML LSCs survival is intrinsic and potentially independent from external factors (Eiring et al., 2015). Even more reassuring is

a report showing a BCR-ABL1 TK independent change in gene expression. It was shown that *MIR10A* is downregulated in CML cells and showed no change in expression after TKI treatment. This led to increased proliferation and cell growth (Agirre et al., 2008).

The knowledge that there are other potential therapeutic targets in CML LSCs has promoted avid research in the field. This has led to the discovery of new compounds that target this malignant cell population. Some of these treatments include PP2A activating drugs (PADs) (Neviani et al., 2013), pioglitazone (Prost et al., 2015), EZH2 inhibitors (EZH2i) (Scott et al., 2016), autophagy inhibitors (Baquero et al., 2018), BETi and MDM2 inhibitors (MDM2i) (Abraham et al., 2016). PADs have successfully been used for inhibiting the high levels of JAK2 activity in TKI treated cells (with inhibition of BCR-ABL1 TK). In this situation, this drug successfully eliminated part of the CML LSCs but failed to completely eradicate the clone (Neviani et al., 2013). Pioglitazones are agonists of PPAR γ and have been shown effective at reducing the expression levels of *STAT5* (Prost et al., 2015). *STAT5* is mainly in its inactive conformation (unphosphorylated) in TKI treated cells but the levels of total *STAT5* remain constant. The unphosphorylated confirmation retains a role in transcriptional regulation in HSCs by displacing other transcription factors from their binding sites (Park et al., 2016). In fact, the use of pioglitazone treatment reduced the expression of *STAT5* and this induced the cells to enter cell cycle and reduced the number of CFCs (Prost et al., 2015). The use of pioglitazone has also been tested in 3 CML patients that were on IM treatment and it was shown effective at inducing complete molecular response (Prost et al., 2015). CML LSCs present a deregulation of the Polycomb Repressive Complex 2 (PRC2) which leads to a downregulation of *EZH1* and an upregulation of *EZH2* (Scott et al., 2016). Targeting *EZH2* with *EZH2i* potentiate the TKI transcriptional signature on CML LSCs and induce them to enter apoptosis while not affecting normal haematopoiesis (Scott et al., 2016). Autophagy is a known mechanism of survival in HSCs and CML LSCs seem to use it for overcoming TKI mediated cell death. The use of autophagy inhibitors induce CML LSCs to enter cell cycle and to differentiate, which sensitises them to TKI treatment (Baquero et al., 2018). Finally, the finding of a deregulated protein network in CML LSCs where MYC and p53 were the main nodes led to the use of BETi and MDM2i in CML LSCs (Abraham et al., 2016). Dual treatment with BETi and MDM2i induced apoptosis and differentiation in CML LSCs.

Although some of these treatment rationales were based on the discovery of certain pathways differentially expressed or with deregulated activity in CML LSCs, and independent of the BCR-ABL1 TK activity (Abraham et al., 2016, Eiring et al., 2015), a

fully comprehensive network of the deregulated pathways in these cells under BCR-ABL1 TK inhibition has not been described. Therefore, a better understanding of the mechanisms underlying CML LSCs persistence would be beneficial for the development of new targeted therapies.

1.3 Aims

The existing literature points towards the existence of a molecular signature in CML that is independent of the TK activity of BCR-ABL1. First, CML LSCs persist TKI treatment despite the drugs successfully inhibiting the BCR-ABL1 TK activity. Second, different signalling pathways seem to be deregulated in CML even during BCR-ABL1 TK inhibition, such as the intrinsic β -catenin pathway, p53 and MYC. However, the existence of a BCR-ABL1 TK independent gene expression signature has not been closely investigated and a more cohesive model would be beneficial for the understanding of CML LSC persistence and biology.

In addition, the current prognosis scores used for CML patients, which are based on clinical factors, lack sensitivity and a big proportion of CML patients that have a poorer outcome are initially scored as low-risk individuals. Therefore, the development of molecular prognostic scores that complement the current EUTOS score may benefit patients that are currently miss-classified and do not benefit from the closer monitoring of high-risk patients.

The main aim of this thesis is to investigate if a gene expression signature independent of the TK activity of BCR-ABL1 actually exists in CML LSCs and if it can be further exploited in the clinical practice, both as a source of new therapeutic targets or as a prognostic biomarker. The main objectives and the plan of investigation have been summarised in the following points:

- I. Definition of a TKI independent (TKIi) signature in chronic phase CML LSCs using whole transcriptome microarray datasets comparing (I) CML and normal HSCs and (II) CML CD34⁺ cells treated and untreated with TKIs.
- II. Investigation of the role of the TKIi signature as a biomarker for disease prognosis both as aggressiveness of the disease and response to IM treatment.
- III. Characterisation of the TKIi signature expression in the different phases of CML.
- IV. Investigation of the effect of drug treatments targeting the TKIi signature in CML CD34⁺ cells.

2 Materials and methods

2.1 Materials

2.1.1 Primary patient material

All samples were collected after written informed consent from the patients. The project had approval from the West of Scotland Research Ethics Committee (REC reference: 15-WS-0077). Samples were processed from peripheral blood or leukapheresis from patients suffering from CML, other haematological malignancies or healthy donors (allogenic haematopoietic stem cell donors). Patients' age, gender and response to IM are summarised in Table 2-1 for CML and in Table 2-2 for non-CML (nCML).

2.1.2 Tissue culture solutions

2.1.2.1 RPMI⁺

- RPMI
- 10% v/v FBS
- 2mM L-glutamine
- 100U/mL penicillin and 100µg/mL streptomycin

2.1.2.2 RPMI⁺⁺

- RPMI
- 20% v/v FBS
- 2mM L-glutamine
- 100U/mL penicillin and 100µg/mL streptomycin

2.1.2.3 DMEM⁺

- DMEM
- 10% v/v FBS
- 2mM L-glutamine
- 100U/mL penicillin and 100µg/mL streptomycin

Table 2-1. Summary of the CML primary patient samples used in this project. The levels of BCR-ABL1 by qPCR refer to the 6 months checkpoint. n/a not available.

Sample	Gender	Age	Disease	BCR-ABL1 > 10%	Additional information
CML444	M	45	CML-CP	No	Used fresh
CML450	M	63	CML-CP	Yes	ELN failure. Used fresh
CML452	M	43	CML-CP	n/a	Used fresh
CML454	M	56	CML-CP	n/a	Used fresh
CML456	F	63	CML-CP	Yes	Change to dasatinib. Used fresh for TKI validation Used from frozen for GO treatment
CML459	n/a	n/a	CML-CP	n/a	Used fresh
CML457	M	60	CML-CP	No	Used fresh
CML469	M	62	CML-CP	n/a	5 days of hydroxyurea before collection Used fresh
CML470	M	69	CML-CP	n/a	Used fresh
MP10AC	n/a	n/a	CML-BC	n/a	Used fresh
CML398	M	61	CML-CP	No	ELN failure. Used from frozen
CML429	n/a	n/a	CML-CP	n/a	Used from frozen
CML441	M	63	CML-CP	n/a	Change to bosutinib after 3 months Used from frozen
CML423	F	27	CML-CP	Yes	ELN failure. Intolerant to IM and NIL Used from frozen
CML460	F	30	CML-CP	n/a	Used from frozen

Table 2-2. Summary of the nCML samples used in this project. n/a not available.

Sample	Gender	Age	Disease	Additional information
PGT160414	M	28	Mantle cell lymphoma	Used fresh
PGT160322	F	49	Lymphoma	Used fresh
PGT160907B	M	62	Lymphoma	Used fresh
PGT170419	M	56	Myeloma	Used fresh
PGT170504A	M	n/a	Lymphoma	Used fresh
PGT170511	F	n/a	Myeloma	Used fresh
nCML035	M	26	Relapsed Hodgkins	Used from frozen
nCML038	M	49	Lymphoma	Used fresh
nCML039	M	20	Allogenic donor	Used fresh
nCML040	F	53	Myeloma/plasmablastic lymphoma	Used fresh
nCML041	M	35	Lymphoma	Used fresh
nCML025	F	n/a	Erwing sarcoma	Used from frozen
nCML029	M	65	Diffuse large B-cell lymphoma	Used from frozen
nCML032	M	48	Lymphoma	Used from frozen
PGT170808	M	54	Myeloma	Used fresh
PGT170830B	F	61	Diffuse large B-cell lymphoma	Used fresh
nCML033	M	34	Allogenic donor	Used fresh
nCML034	F	28	Allogenic donor	Used fresh
PGT170816	F	n/a	Lymphoma	Used fresh
nCML042	M	68	Lymphoma	Used fresh

2.1.2.4 Serum free media (SFM)

- IMDM
- 2mM L-glutamine
- 100U/mL penicillin and 100µg/mL streptomycin
- 100µM 2-mercaptoethanol
- 20% v/v bovine serum albumin, insulin, and transferrin (BIT)

2.1.2.5 Physiological growth factors (PGF)

- 0.2ng/mL SCF
- 1ng/mL G-CSF
- 0.2ng/mL GM-CSF
- 1ng/mL IL6
- 0.05ng/mL LIF
- 0.2ng/mL MIP1 α

2.1.2.6 FACS solution

- PBS
- 2% v/v FBS

2.1.2.7 DAMP

- 50U/mL of DNase
- 2.5mM MgCl₂
- 14mM trisodium citrate
- 1% v/v human serum albumin (HSA)
- PBS

2.1.2.8 Magnetic activated cell sorting buffer

- PBS
- 2mM EDTA
- 1% v/v HSA

2.1.2.9 Gemtuzumab-ozogamicin (GO)

- Sterile water
- 1mg/mL GO

2.1.2.10 Imatinib (IM)

- Sterile water
- 100mM IM

2.1.3 Flow cytometry

All flow cytometry data collection was performed in a FACSCanto (BD, Oxford, UK). Flow cytometry data was analysed using FlowJo 10 (FlowJo LLC, Ashland, Oregon, USA).

2.1.3.1 Antibodies

All antibodies were purchased from BD (Oxford, UK).

Epitope	Clone
Kl67	B56
γH2AX	N1-431
CD34	581
CD33	P67.6

2.1.4 Molecular biology

Retrotranscription, pre-amplification and digestions were performed in a Techne TC-412 thermocycler. A 7900HT Fast Real-Time PCR System (Life Technologies) was used for quantitative PCR (qPCR). Microfluidics qPCR was performed using Fluidigm's Biomark™ HD system.

2.2 Methods

2.2.1 *Cell biology*

2.2.1.1 **Cell counts**

CD34⁺ and mononuclear cells were counted by trypan blue dye exclusion using a haemocytometer. Cell lines were counted using trypan blue dye exclusion using the EVE™ Automated Cell Counter (NanoEnTek) that also allowed measurement of the average size of the cells.

2.2.1.2 **Recovery of frozen primary cells**

Frozen cells were thawed rapidly at room temperature immediately after removing them from liquid nitrogen. On thawing, the content of the vials was added to a 50mL sterile centrifuge tube. 10mL of pre-warmed DAMP (20°C) were added drop-wise to the cells over 20 minutes providing continuous agitation. Cells were then centrifuged at 200g for 10 minutes and the supernatant and any cell debris aggregates were discarded. This process was repeated another two times. Cells were counted and resuspended in SFM+PGF at a density of 1×10^6 cells/mL (CD34⁺ cells) or 1×10^7 cells/mL (MNC).

2.2.1.3 **Magnetic activated cell sorting (MACS)**

Mononuclear cells (MNC) were enriched in CD34⁺ cells using MACS. The cells were centrifuged at 300g for 10 minutes and the supernatant was discarded. The pellet was resuspended in 300µL of room temperature MACS buffer per million CD34⁺ cells (expected number in the sample based on cell counts and previous CD34 fluorescent labelling analysed by flow cytometry). After that, 30µL of CliniMACS anti-CD34 magnetic bead-conjugated antibodies per million CD34⁺ cells were added to the suspension and mixed gently. The cells and the beads were incubated for 30 minutes at room temperature. Once the incubation was finished, 5mL of MACS buffer per million CD34⁺ cells was added to the suspension and centrifuged 10 minutes at 300g. Supernatant was discarded and cells were resuspended in 500µL of MACS buffer per million CD34⁺ cells and applied to a LS column previously primed with 3mL of MACS buffer. The column flow-through was collected in a sterile 50mL centrifuge tube. Once the flow-through stopped, 3mL of MACS buffer were added on top of the column. This last step was repeated 2 more times. This flow-through contained the CD34⁺ population. In order to get the CD34⁺ cells 5mL of MACS buffer were added on top of the column and pushed firmly with the plunger through the column. This flow-through was collected in a new sterile 50mL centrifuge tube. Purity of CD34⁺ cells was confirmed in both fractions and

cells were counted. Most of the samples were enriched for CD34⁺ cells and cryopreserved where appropriate by Dr Alan Hair.

2.2.1.4 Surface antigens staining and detection

Detection and analysis of surface antigens was performed with fluorescently labelled antibodies and detected by flow cytometry. An aliquot of cells ($\approx 20,000$) was transferred to a flow cytometry tube and centrifuged for 5 minutes at 350g. Supernatant was discarded and 5 μ L of anti-CD34 APC and 5 μ L of anti-CD33 PE-Cy7 antibodies were added to the cell. Additional tubes with (I) no antibodies, (II) only anti-CD34 APC, (III) only anti-CD33 PE-Cy7 and (IV) the isotype control for PE-Cy7 were prepared. The antibodies were incubated for 20 to 30 minutes at 4°C in the dark. After the incubation the cells were washed with 700 μ L of FACS solution and centrifuged at 350g for 5 minutes. The supernatant was discarded and the cells were analysed in a FACSCanto cell analyser.

2.2.1.5 Culture of CML CD34⁺ cells for validation of the TKli genes

For these experiments the CML CD34⁺ cells were cultured immediately after CD34⁺ enrichment (>90% purity). Cells were not frozen at any time. After counting the cells using a haemocytometer, cells were washed and resuspended in SFM at a density of 1×10^6 cells/mL. Cells were then seeded in a 6-well plate and grouped as no drug control (NDC) or IM treated. IM was added at a final concentration of 5 μ M. Cells were cultured at 37°C and 5% CO₂ for 7 days. IM was added again at day 4 without washing the cells. On day 7 cells were sorted by FACS by Ms Jennifer Cassels for viable cells (DAPI⁻) and used for RNA extraction.

2.2.1.6 Culture of primary cells for drug response studies

After recovery cells were resuspended at a density of 2×10^5 cells/mL in SFM+PGF in the presence or absence of 2 μ M IM and/or different concentrations of GO (10, 30, 100, 300 and 1000ng/mL) and cultured at 37°C and 5% CO₂. Treatment was delivered in three different regimens:

- 72 hours of IM+GO simultaneous combination.
- 72 hours IM followed by 72 hours of GO.
- 72 hours of GO followed by 72 hours of IM.

Following 72h the cells were washed in SFM and centrifuged for 10 minutes at 300g three times in order to eliminate the first drug. After treatment the cells were counted and used for downstream experiments.

2.2.1.7 Assessment of apoptosis

Induction of apoptosis was measured using flow cytometry. The cells ($\approx 100,000$) were washed in Hank's Balanced Salt Solution (HBSS) and centrifuged at 350g for 5 minutes. Supernatant was discarded and a solution of 98 μ L of HBSS, 2 μ L of Annexin V and 0.1 μ L of 1mg/mL DAPI were added and incubated for 15 minutes at room temperature in the dark. Once incubation was finished the cells were analysed in the FACSCanto cell analyser. During apoptosis and necrosis the phosphatidylserine in the cell membrane flips towards the outer part of the plasma membrane allowing Annexin V, a protein with high affinity for phosphatidylserine, to bind (Koopman et al., 1994). By conjugating Annexin V to a fluorescent dye and using a vital dye, such as DAPI, it is possible to discriminate between viable cells (Annexin V⁻DAPI⁻), early apoptotic cells (Annexin V⁺DAPI⁻) and late apoptotic or necrotic cells (Annexin V⁺DAPI⁺), as shown in Figure 2-1 (Vermes et al., 1995).

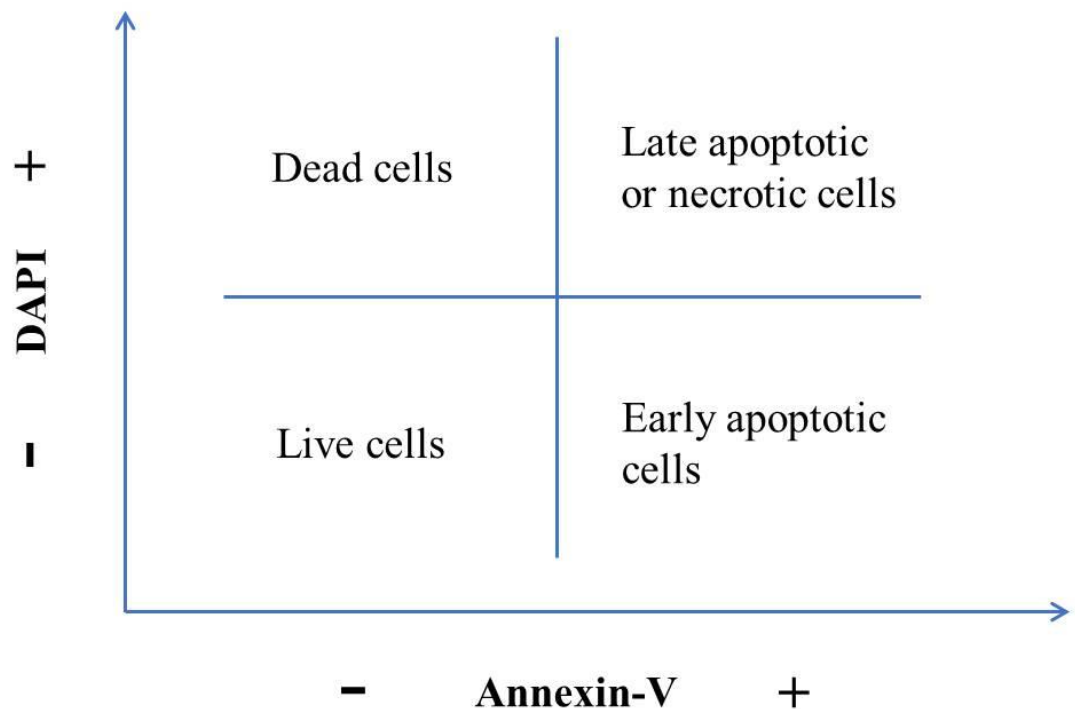


Figure 2-1. Classification of the cells in viable, apoptotic and necrotic by AnnexinV and DAPI staining. AnnexinV⁻ DAPI⁻ cells are classified as viable cells. AnnexinV⁺ DAPI⁻ cells are classified as early apoptotic and AnnexinV⁺ DAPI⁺ cells are classified as late apoptotic or necrotic cells.

2.2.1.8 Staining and detection of intracellular antigens

Intracellular protein expression was also measured using flow cytometry. For this, the cells need to be fixed and the plasma membrane permeabilised in order for the antibodies to reach the epitopes that they recognise. This was performed using BD Cytofix/Cytoperm™ Fixation/Permeabilisation Kit. Cells ($\approx 150,000$) were transferred to flow cytometry tubes and washed with FACS solution and centrifuged at 350g for 5 minutes twice. The cell pellet was vortexed for reducing cell aggregation and resuspended in 250 μ L of fixation/permeabilisation solution and incubated for 20 minutes at 4°C. Cells were washed in perm/wash™ buffer (i.e. centrifuged at 350g for 5 minutes) then resuspended in 400 μ L of FACS solution. Tubes were sealed with parafilm and stored at 4°C until further use.

On day of analysis, cells were centrifuged at 350g for 5 minutes to remove FACS solution. Plasma membrane was further permeabilised by incubating the cells in perm/wash™ buffer for 15 minutes. Cells were centrifuged again at 350g for 5 minutes and buffer was discarded. 5 μ L of anti- γ H2AX antibody were added to the cell pellet and incubated for 20 minutes at 4°C in the dark. Afterwards the cells were washed twice with PBS (centrifuged at 350g for 5 minutes) and incubated at room temperature with 15 μ L of anti-KI67 antibody in the dark. Cells were washed one more time in PBS (centrifuged at 350g for 5 minutes) and resuspended in 200 μ L of a dilution of 1:200 of DRAQ7 in PBS for DNA staining. The cells were incubated for additional 15 minutes at room temperature in the dark and analysed on the FACSCanto.

2.2.1.9 Colony forming cell (CFC) assay

After 72h or 144h treatment 3,000 cells from the selected wells were harvested and mixed with 3mL of Methocult® H4034. This mix was then split evenly between two 35mm² plates covering the entire surface of the plates. All the plates from each sample were placed inside a 23.5cm² plate and 2 plates containing just water were added to avoid the Methocult drying. Cells were cultured for no less than 9 days at 37°C and 5% CO₂. Then, colonies were counted and scored as erythrocytes (E), granulocytes (G), macrophages (M), granulocyte-macrophage (GM) or granulocyte-erythrocyte-macrophage-megakaryocyte (GEMM) using the Stem Cell Technologies manual as reference.

2.2.1.10 Long-term culture

2.2.1.10.1 Preparation of the feeder cells layer

A mix of 7.5×10^4 M210B4 and 7.5×10^4 SI/SI cells was seeded in each well with 500 μ L of RPMI⁺ and 500 μ L of DMEM⁺ in a 24-well nunc® plate and were incubated for 24 hours at 37°C and 5% CO₂. After 24 hours visual inspection of the cells confirmed their adhesion to

the plastic. If the cells had adhered to the plastic, 800µL of media were removed from each well and 10µL of 1mg/mL mitomycin C was added to the cells in order to get a final concentration of 50µg/mL. The cells were incubated at 37°C and 5% CO₂ for 30 minutes and then, using a Pasteur pipette each well was washed 3 times with sterile PBS. After washing 1mL of Myelocult H5100 supplemented with 1µM hydrocortisone was carefully added to each well, making sure the cells were not dislodged. The addition of mitomycin C stops the cells from dividing, therefore, keeps a constant number of live feeder cells over the duration of the experiment. Perimeter wells were filled with sterile water in order to reduce loss by evaporation.

2.2.1.10.2 Seeding and culture of primary cells on feeder cells

After completion of drug treatment (72h or 144h) 50,000 cells from the each of the wells of interest (those with 0, 30 and 100ng/mL of GO±IM) were transferred to the plates containing the feeder cells. The cells were placed in an incubator at 37°C and 5% CO₂ for six weeks.

Every week half of the media was replaced using filtered pipette tips. For wells placed at the edge of the plate a 100µL surplus of Myelocult supplemented with hydrocortisone was added to counter the effects of evaporation.

2.2.1.10.3 Harvest

After six weeks the media was collected into a 15mL centrifuge tube (collection tube) and the wells rinsed gently twice with 1mL of sterile PBS to remove the remaining media. After each rinse the PBS was added into the collection tube. After the two rinses, 200µL of 0.25% w/v trypsin were added into each well and the cells were placed for 15 minutes in an incubator at 37°C and 5% CO₂. Immediately afterwards 40µL of filtered FBS was added into the wells to neutralise the trypsin while dislodging the cell layer by pipetting. The cells were resuspended in 1mL of 2% FBS IMDM and transferred to the collection tubes. Each well was rinsed two more times with 1mL of 2% FBS IMDM and contents transferred to the collection tube. The cells were centrifuged at 276g for 7 minutes and the supernatant discarded. This process was repeated another two times resuspending the cells in 3mL of 2% FBS IMDM. Any remaining IMDM was carefully aspirated with a pipette without disturbing the pellet. The volume of the remaining IMDM was measured and it was used for resuspending the cells. The number of cells was counted using an haemocytometer and trypan blue dye exclusion. The cells were finally transferred to 3mL of Methocult and mixed, plated, cultured and counted as described for the CFC assay (2.2.1.9).

2.2.1.11 Cell lines drug response analysis

K562 CML myeloid blast crisis cell line was cultured in RPMI⁺ and BV173 CML lymphoid cell line was cultured in RPMI⁺⁺. Both cell lines were maintained between 2×10^5 and 1×10^6 cells/mL in 75cm² cell culture flask and incubated at 37°C and 5% CO₂. Drug treatments were performed in 24 or 96 wells plates at a cell density of 2×10^5 . The absence of mycoplasma was confirmed using Lonza Mycoalert detection kit. Cell line composition (i.e. identity, genotype) was not confirmed once obtained from the Paul O’Gorman Leukaemia Research Institute biobank.

2.2.1.12 Resazurin

Synergy studies in cell lines were performed using cell viability reported by resazurin. Resazurin is a non-fluorescent blue dye. However, when reduced by cell metabolism it is transformed into resorufin, which is a fluorescent red dye that can be excited at 530nm and emits at 590nm. As no previous steps are needed for its use, resazurin is a reasonable approach for high throughput analysis of cell viability (Ansar Ahmed et al., 1994). After treatment, 10µL of 500µM resazurin were added to 90µL of cell suspension and incubated 4 hours at 37°C and 5% CO₂. Fluorescence was measured in a spectrophotometer at 590nm after stimulating the dye at 535nm using wells with only culture media (no cells) as blank and non-treated cells as control. Effect of the each drug combination was calculated as the ratio of the fluorescence emission of each condition compared with the NDC. The blank was calculated adding resazurin to a well containing no cells and therefore, no conversion to resorufin would be observed. All the values were subtracted the blank before performing any calculation.

2.2.2 Molecular biology

2.2.2.1 RNA extraction

Total RNA extraction of cell lines was performed using Qiagen RNeasy Mini kit. Total RNA from patient’s samples was performed using Qiagen RNeasy Micro kit or Arcturus PicoPure depending on the sample size of the condition with the lowest number of cells. Extraction was performed following manufacturer’s protocol. Total RNA was quantified in a spectrophotometer at 260nm. The samples used for RNAseq were quantified and the RNA integrity numbers (RIN) were assessed on an Agilent 2100 bioanalyser system. RNA was kept at -80°C for long-term storage.

2.2.2.2 Primer design

Primers were designed using Primer Blast (Ye et al., 2012) selecting melting temperatures close to 60°C (±1) and with amplicon sizes between 80 and 150bp (MIR21 amplicon had a

length of 50bp). When the gene had more than one exon, primers were designed to cover an exon-exon junction. Primers were synthesised by IDT (Leuven, Belgium). Sequences can be found in Appendix I.

2.2.2.3 Retrotranscription

RNA was reverse transcribed into cDNA using high capacity cDNA reverse transcription kit (Life Technologies) following the manufacturer's protocol. cDNA was kept at -20°C for long-term storage.

2.2.2.4 Quantitative polymerase chain reaction (qPCR)

PCR duplicates the number of molecules of DNA with the targeted sequence on each cycle of the reaction in the presence of DNA polymerase, dNTPs (deoxynucleotides triphosphate), primers flanking the sequence of interest and salt (MgCl₂). Each cycle is composed of the following phases: denaturation, annealing, and extension. During denaturation the two strands of DNA separate allowing the primers to bind during the annealing phase. During extension the DNA polymerase add the dNTPs complementary to the model strand to the novel strand of DNA.

SYBR Green is a molecule that when bound to double strand DNA emits fluorescence at 524nm (Zipper et al., 2004). This allows quantification of the amount of DNA in a particular sample by emitted fluorescence. By detecting the amplification cycle at which the fluorescence intensity reaches the set intensity threshold (Ct or threshold cycle) it is possible to determine the amount of original DNA copies of the sequence of interest.

Each qPCR experiment was performed for 40 cycles using PowerUp SYBR Green master mix (Life Technologies), 10ng of cDNA and 500nM of each of the primers. Activation of the polymerase required heating the reaction for 2 minutes at 50°C and another 2 minutes at 95°C. Each of the following reaction cycles had a 15 second denaturation step at 95°C and 60 seconds annealing and extension step at 60°C. Dissociation curves were performed for all reactions and both non template control (NTC) and no DNA control (RNA not retrotranscribed) reactions were performed for every set of primers and experiment.

2.2.2.5 Fluidigm

The Fluidigm platform is a microfluidics qPCR system that allows performing multiple reactions every run. During this project, a 48.48 chip was used, which amplifies 48 genes on 48 samples, making a total of 2,304 reactions.

The cDNA of each sample was pre-amplified for 18 cycles using the PCR multiplex PCR kit (Qiagen). Each reaction contained a pool of all the primers of interest at 50nM

concentration and a maximum 12.5ng of cDNA (some samples yielded very low concentrations of RNA and higher concentrations were not possible). The polymerase was activated at 95°C for 15 minutes and each cycle comprised of 30 seconds of denaturation at 94°C, 90 seconds of annealing at 60°C and 60 seconds of extension at 72°C. A final extension of 30 minutes at 72°C was performed. The samples were treated with 0.5U/μL of exonuclease I (New England Biolabs, Ipswich, MA, USA) for 30 minutes at 37°C. The enzyme was inactivated at 80°C for 15 minutes. This step eliminates the non-incorporated primers avoiding non-specific amplification during the qPCR. The samples were diluted 1:5 and stored.

The 48.48 chip was primed with control line fluid in the IFC controller MX and each of the primer wells was filled with a 5μL solution containing 1X assay loading reagent, DNA suspension buffer and 5μM of each of the primers of the pair assigned to the well. Each sample was loaded with 5μL of 1X SsoFast™ EvaGreen Supermix with low ROX (Bio-Rad), 1X DNA binding dye sample loading reagent (Fluidigm) and 45% v/v of the pre-amplified cDNA assigned to the well. The reaction in the Biomark activated the enzyme at 95°C for 1 minute and performed 30 cycles of denaturation at 96°C for 5 seconds and annealing and extension at 60°C for 20 seconds. A melting curve was generated at the end of the qPCR for every reaction.

2.2.2.6 RNA sequencing

RNA for RNAseq was reverse transcribed using the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara, Saint-Germain-en-Laye) by Glasgow Polyomics. cDNA library preparation for RNAseq was performed using Nextera library preparation kit by Glasgow Polyomics. RNA sequencing was performed using Illumina HiSeq and NextSeq sequencers. Sequencing reads of both instruments were merged by Glasgow Polyomics.

Illumina platforms utilise an imaging system to identify the sequence of each of the cDNA molecules (Metzker, 2009). To facilitate the detection, Solid-phase or bridge PCR generates clusters of each of the cDNA fragments on a surface. Once the cluster is generated, primers complimentary to the library adaptors are hybridised to the cDNA fragments. Then, a mix of fluorescently labelled nucleotides and DNA polymerase are released to the reaction chamber and a single nucleotide is added to the sequence. The addition of a blocker and the fluorescent label prevents subsequent nucleotides from being added to the sequence. After washing the excess nucleotides, the imaging system detects the fluorescent signal from each cluster and identifies the nucleotide with which it is

associated. Then the fluorescent label and the blocker get cleaved from the nucleotide and a new cycle starts until the sequence is complete.

The reads generated in the HiSeq were pair-ended and 75 base pairs long while the reads generated in the NextSeq platform were single-ended and 75 base pairs long.

2.2.3 Bioinformatics

Data analysis was performed using R 3.5.0 running under macOS 10.13.6 unless otherwise stated.

2.2.3.1 Microarray analysis

Data generated using Affymetrix HuGe 1.0 ST microarray were imported into R using *oligo* (Carvalho and Irizarry, 2010) and pre-processed using the robust multichip average (RMA) method (Irizarry et al., 2003), which is a quantile normalisation approach. Depending on the analysis, probe set values were summarised to either the gene level or RefSeq transcript level.

Data generated by other chip types (Affymetrix Human Genome U133 Plus 2.0 and Affymetrix Human Genome U133A 2.0) was imported into R using *affy* (Gautier et al., 2004). Pre-processing was performed by RMA and the probe set intensity values were summarised at the RefSeq level.

Differential gene expression was calculated using *limma* (Smyth, 2004, Ritchie et al., 2015). *limma* uses empirical Bayes linear models for estimating the differential gene expression. This takes in account the gene expression differences between the different groups for each gene but it also takes in account the overall gene expression differences for computing the empirical Bayes moderated t-statistics and the associated p-value. It also allows analysing experiments with complex designs, such as those with multiple groups and batch/individual effect. Technical replicates were accounted using the *duplicateCorrelation* function and type I error was controlled by using Benjamini-Hochberg (BH) correction (Benjamini and Hochberg, 1995). Specific thresholds for statistical significance are discussed in each chapter.

Statistically significant probe sets were annotated to gene names (HGNC symbols) or RefSeq accession numbers with *biomaRt* (Durinck et al., 2009) connecting to the Ensembl release December 2014 (Cunningham et al., 2015).

2.2.3.2 *In silico* validation of gene lists

Over and under representation analyses were performed using both the hypergeometric test and Monte-Carlo sampling (Metropolis and Ulam, 1949). The universe, that is, the identity of all the elements present in the analysis, was defined according to the technical limitations (i.e. genes represented by the probe sets of a particular microarray chip or transcripts detected by RNAseq). Monte-Carlo functions were written in-house while hypergeometric distribution calculations were based on the base R function *dhyper*.

A hypergeometric distribution is often described as the distribution of probabilities of drawing a number of white balls from a known mixture of white and black balls without replacement. This can be used for overrepresentation analyses by calculating the cumulative probability of drawing the same or higher number of white balls, which would equate to the p-value. For underrepresentation analyses the cumulative probability of drawing the same or lower number of white balls is calculated instead. For example, it could be used to calculate if a list of 100 differentially expressed genes between normal and leukaemic samples have a higher proportion of cancer related genes than what would be expected in a random set of 100 genes in the same universe.

Monte Carlo simulation can also be used to assess the significance of analysis results. Here, the same analysis is run multiple times with the data randomised to provide a null distribution over the analysis result. For example, random sets of data can be generated by permuting the sample or gene names in a microarray or RNAseq dataset (the permuted values are specified when describing each experiment). This maintains a similar distribution to that of the original data resembling best the original analysis. This allows calculating the probability of obtaining a value equal or higher than that obtained with the real data by dividing the number of iterations that generated a value equal or higher to the real one by the total number of iterations (overrepresentation). For underrepresentation analysis the number of permutations with equal or lower value than the real data will be divided by the number of total permutations. This ratio of permutations of interest by total permutations would equal to the p-value of the test.

$$p = \sigma_{v \geq x}(k)/n$$

$$p = \sigma_{v \leq x}(k)/n$$

Equation 1. Calculation of the p-value for Monte Carlo simulations. For overrepresentation analysis (up) the p-value equates to the number of iterations (k) with a value (v) equal or greater than the real value (x) divided by the total number of iterations (n). In underrepresentation analysis (down) v should be equal or smaller than x.

2.2.3.3 Principal components analysis (PCA)

PCA is a common technique used for data exploration. It allows identifying batch or patient effects and can provide with a quick insight of the main factors contributing to the variance. This variance is assessed by the generation of eigen vectors (vectors of length 1) with a number of components equal to the number of observations (i.e. number of genes in the universe). The eigen vector which product with each of the samples/samples has the highest variance is selected as principal component 1 (PC1). Then the process is repeated with the vectors perpendicular to PC1 from which the one with the highest variance is selected as PC2. The process is repeated for subsequent PCs. PCA was performed using *prcomp* function from base R.

Association of the PCs with each factor of interest was performed using Kruskal-Wallis test on the values of each eigen vector divided on groups based on the levels of each factor. The script for this calculation was kindly provided by Dr Lisa Hopcroft and later modified for adapting for the data format used in the different experiments.

2.2.3.4 Generation of a classifier: Support vector machine (SVM)

A common approach for classifying a binary factor (e.g. CML/normal, responder/non-responder) would be the use of a hyperlane. A hyperlane is an element with one less dimension than the space it is trying to classify (e.g. a line when classifying elements in a plane). However, using a hyperlane directly on the data can fail to classify non-linear data. To prevent this, the dimensions of the data can be increased using a kernel and then, applying the hyperlane based on the training dataset (James et al., 2013). In this thesis, the SVM classifier generated used the scikit-learn SVM model with default options (RBF kernel) (Pedregosa et al., 2011). This was kindly performed by Dr Simon Rogers using Python programming language under macOS.

2.2.3.5 qPCR analysis

Relative expression of the test genes was calculated by subtracting the mean of the Ct values of the reference genes (*ENOX2*, *GAPDH*, *RNF20*, and *TYW1*) to the Ct value of the test gene within each sample (ΔCt). These ΔCt values were used as normalised gene expression values and differential gene expression was calculated using *limma* (Ritchie et al., 2015). Genes were considered to be differentially expressed when the BH-adjusted p-value was lower than 0.05. The confidence interval of the $\Delta\Delta Ct$ ($\log_2$ fold change) was also calculated by *limma*.

2.2.3.6 RNAseq analysis

Quality of sequencing was assessed using *FastQC* (Andrews, 2010), and sequences were trimmed using *TrimGalore* (Krueger, 2015) with 3' end trimming and paired parameters, using a quality score threshold of 20. The reference genome (Homo_sapiens.GRCh38.dna.primary_assembly.fa) was obtained via Ensembl on 20th March 2018 and reads were aligned to the genome using *hisat2* (Kim et al. 2015). Count matrix was generated using *featureCounts* (Liao et al. 2013). This pre-processing was kindly performed by Ms Joana Bittencourt-Silvestre.

Differential gene expression was calculated using *DESeq2* (Love et al., 2014). *DESeq2* uses generalised linear models for the analysis of differential gene expression in RNAseq count matrixes. This provides flexibility for the analysis of different experimental designs. This package also uses empirical Bayes shrinkage of the dispersion and the fold changes to reduce the effect of very variable genes and genes with very low counts. Additionally, to further reduce the noise introduced genes with low counts, only the genes with more than 100 total counts (among all the samples) were selected for differential gene expression analysis. Type I error was controlled by using Benjamini-Hochberg (BH) correction (Benjamini and Hochberg, 1995) and all genes with a q-value equal or smaller than 0.1 were considered significantly differentially expressed.

2.2.3.7 Pathway overrepresentation analysis

Gene lists were analysed for overrepresentation of pathways and gene ontology terms (GO-terms) (Ashburner et al., 2000, The Gene Ontology Consortium, 2017) using Protein ANalysis THrough Evolutionary Relationships (PANTHER) (Mi et al., 2017) and Consensus Path DB (CPDB) (Kamburov et al., 2013). PANTHER possesses its own curated biological pathway database and calculates overrepresentation and underrepresentation (Fisher's exact test) of the members of each pathway using the PANTHER annotated proteins as universe. CPDB is associated with a number of biological pathways, protein complexes, drug interaction and GO-terms datasets. In this project CPDB has been used for interrogating Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017), Reactome (Fabregat et al., 2018), BioCarta (www.biocarta.com) and GO-terms at level 3 of the GO hierarchy. Over and under representation were calculated using the hypergeometric test and corrected for multiple testing using BH correction (Benjamini and Hochberg, 1995) using the list of genes analysed in the experiment (detected genes in RNAseq or genes with assigned probe sets in microarray experiments) as the universe. A threshold of 3 or more member of the pathway or GO-term was set for avoiding singletons.

2.2.3.8 Drugs synergy calculation

When used together, different drugs and compounds can have the same effect than the sum of their individual effects (expected effect), in which case it is said that they have independent effects. However, this is not always the case and, when used together, different compounds can have a reduced (i.e. the compounds are antagonistic) or enhanced effect (i.e. the compounds are synergistic). For the purpose of identifying drug interactions in this thesis, combination effect were compared with expected effects using Bliss equation (Bliss, 1939) as shown in Equation 2. Statistical significance was calculated using Student T-test.

$$Bliss\ coefficient = Effect_{observed} - (Effect_{DrugA} + Effect_{DrugB} - Effect_{DrugA} \times Effect_{DrugB})$$

Equation 2. Bliss coefficient is the difference of effect between the observed and expected effects of the combination of two drugs. The coefficient is given in unit ratio and allows identifying synergistic and antagonistic effects.

3 Results (I): Identification of a TKI independent signature in CML LSCs

3.1 Introduction

The presence of the BCR-ABL1 fusion protein and its coding gene has been long associated with CML (Rowley, 1973, Konopka et al., 1984, Shtivelman et al., 1985). The presence of this tyrosine kinase in all CML cells lead to the development of a small molecule inhibitor, imatinib, which is able to bind to the catalytic pocket of the kinase, preventing the binding of ATP and, therefore, inhibiting the tyrosine kinase activity of BCR-ABL1 (Carroll et al., 1997, Druker et al., 2001a). This activity named this family of molecules as tyrosine kinase inhibitors (TKIs). However, a primitive quiescent cell population (known as leukaemic stem cells or LSCs) evades apoptosis despite TKI treatment (Graham et al., 2002), representing a reservoir of leukaemic cells that persist to reinitiate the disease upon treatment withdrawal (Holyoake et al., 2001). Although different dosing strategies have been studied for the eradication of this cell population, it persists and reinitiates the disease after treatment discontinuation in most patients even when *BCR-ABL1* transcripts are not detectable by qPCR (Clark et al., 2017, Ross et al., 2013, Mahon et al., 2010). Both their capacity to persist TKI treatment and to reinitiate the disease highlight the key relevance of eradicating LSCs in the pursuit of a cure for CML.

A closer analysis of the molecular mechanism involved in the eradication of CML progenitor cells ($CD34^+CD38^+$) suggests that not only the inhibition of BCR-ABL1 TK but also the inhibition of KIT is essential for their eradication (Corbin et al., 2013). However, the same study was not able to link KIT activity with LSCs persistence and suggested that another pathway may be conferring survival capabilities during TKI treatment. Research performed in Glasgow confirmed that CML LSCs do not require BCR-ABL1 TK signalling for survival and propose that CML LSCs overcome oncogene addiction through different pathways (Hamilton et al., 2012). Recent publications show that p53, MYC (Abraham et al., 2016) and EZH2 pathways (Scott et al., 2016) are deregulated in CML LSCs and targeting them has a potent effect in the elimination of CML LSCs.

The absence of detectable oncogene addiction (Hamilton et al., 2012, Corbin et al., 2013) and the growth advantage observed in CML LSCs (Cashman et al., 1998) has motivated my own investigations into the molecular mechanisms driving the CML LSCs phenotype. The aim of this chapter is to investigate if there is a gene expression signature in CML that is not dependent on the TK activity of BCR-ABL1. To do so, it was assumed that genes whose expression was not affected by TKI treatment were not governed by the TK activity of BCR-ABL1. In order to find that signature, three different microarray datasets were

analysed to compare (I) CML and normal stem cells and (II) CML progenitor cells before and after TKI treatment (Figure 3-1). This identified a list of genes differentially expressed in CML compared with normal controls that were not affected by TKI treatment. In order to confirm the gene expression signature, expression changes were validated by qPCR.

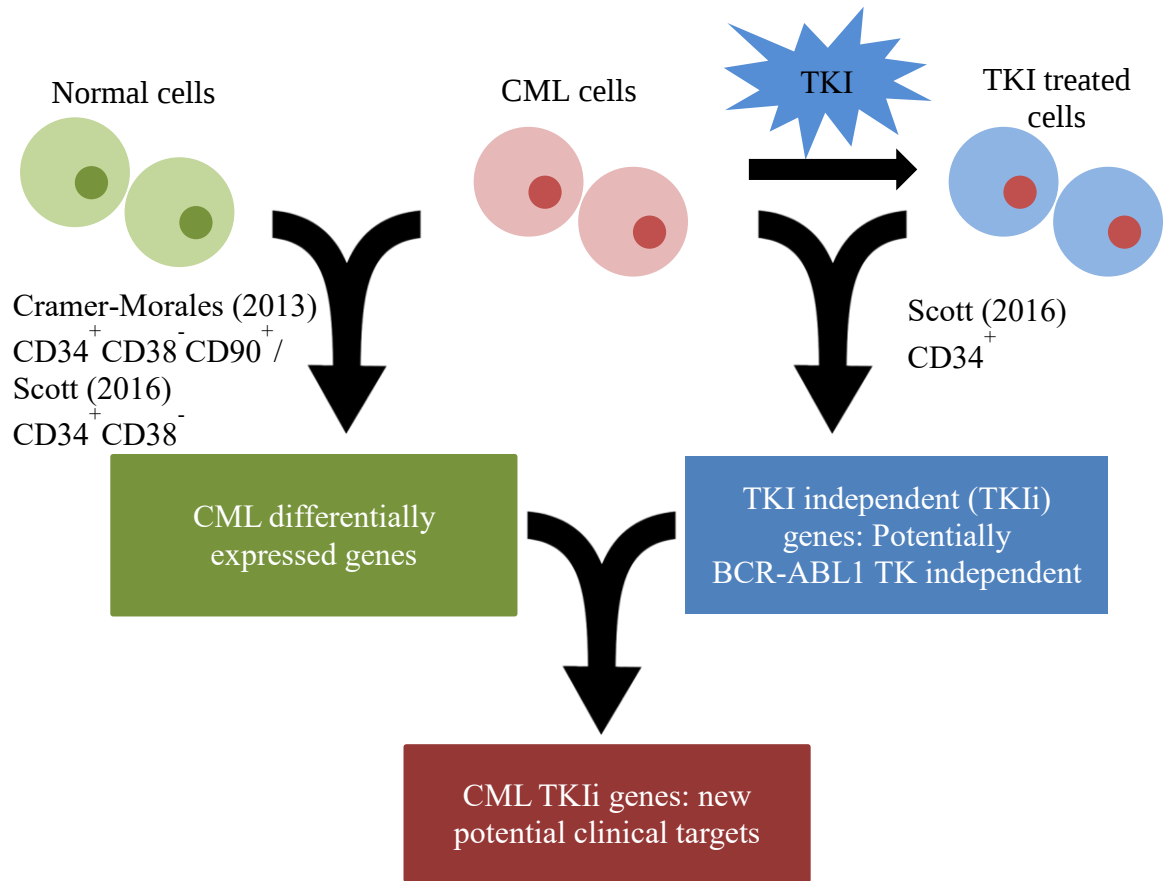


Figure 3-1. Summary of the microarray analysis performed in this chapter for defining the TKI independent (TKIi) transcriptomic signature.

3.2 Identification of the TKI signature using transcriptomic datasets

3.2.1 *Description of the microarray datasets*

3.2.1.1 **Dataset TKIFP: CML LSCs' transcriptional response to TKI**

In the TKIFP dataset (Scott et al., 2016) gene expression levels from LSCs of six different patients were measured using Affymetrix HuGe 1.0 ST chips. Gene expression was measured at baseline (0h) and after eight hours of treatment (8h) for CD34⁺CD38⁻ and after 7 days treatment (7d) for CD34⁺ cells. The cells were treated with 5μM imatinib, 150nM dasatinib or 5μM nilotinib and a second dose of the drug at the same concentration was applied during the 4th day of treatment. The cells were sorted again for live cells after 7d of treatment. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Manchester, UK). Baseline time point was performed using technical triplicates while the replicates of the other time points were performed with each of the different drugs. For the objective of this chapter only the 0h and the 7d data were analysed as preliminary analysis of the 8h time point revealed little effect of the TKIs on the gene expression. The first two principal components (accounting for 50% of the total variability in these data; Figure 3-2A) show samples clustering according to both patient and the presence of TKI (Figure 3-2B), indicating that the differential expression analysis needs to control for each patient sample and a paired analysis approach should be adopted.

3.2.1.2 **Dataset CMLDV: transcriptional differences between CML and normal**

For the CMLDV dataset (Abraham et al., 2016) cells from chronic phase CML patients (n=3) and normal bone marrow donors (n=3) were sorted for CD34⁺CD38⁻ (HSC/LSC) and for CD34⁺CD38⁺ (progenitor cells: HPC/LPC). Total RNA was extracted using RNeasy Micro Kit (Qiagen) when the number of isolated cells was less than 5x10⁵, and RNeasy Mini Kit (Qiagen) when the number of isolated cells was between 5x10⁵ and 1x10⁷. Gene expression was measured using Affymetrix HuGe 1.0 ST microarray chips performing technical duplicates (Array Express accession number E-MTAB-2581). Only the HSC/LSC cells were included in the analysis as they were the most closely matching population to our study population (LSCs). Principal components analysis (PCA) showed normal HSCs closely clustered together while CML LSCs seemed to have a more heterogeneous transcriptome (Figure 3-2D). Technical duplicates clustered tightly together.

3.2.1.3 Dataset CMLMC: transcriptional differences between CML and normal

For the generation of the CMLMC dataset (Cramer-Morales et al., 2013) five different cell populations of varying maturity—HSC (CD34⁺CD38⁻CD90⁺), MPP (CD34⁺CD38⁺CD90⁻), CMP (CD34⁺CD38⁺CD123⁺CD45RA⁺), GMP (CD34⁺CD38⁺CD123⁺CD45RA^{low}) and MEP (CD34⁺CD38⁺CD123⁻CD45RA⁻)—were sorted from healthy donors (n=3), chronic phase (n=6), accelerated phase (n=4) and blast crisis CML patients (n=2). Total RNA was extracted using RNeasy Micro Kit or RNeasy Mini Kit (Qiagen), depending on the number of cells. Gene expression was measured using Affymetrix HuGe 1.0 ST microarray chips. Normalised and quality controlled expression values were downloaded from Stemformatics (Wells et al., 2013) (<http://www.stemformatics.org/>, Gene Expression Omnibus accession number GSE47927). In this chapter only the HSC data from normal and chronic phase patients was used as it is the most primitive population and is the best match to our study population (LSCs). PCA demonstrated that there were differences between normal and CML and that each condition (i.e. CML/normal) forms its own particular cluster (Figure 3-2F).

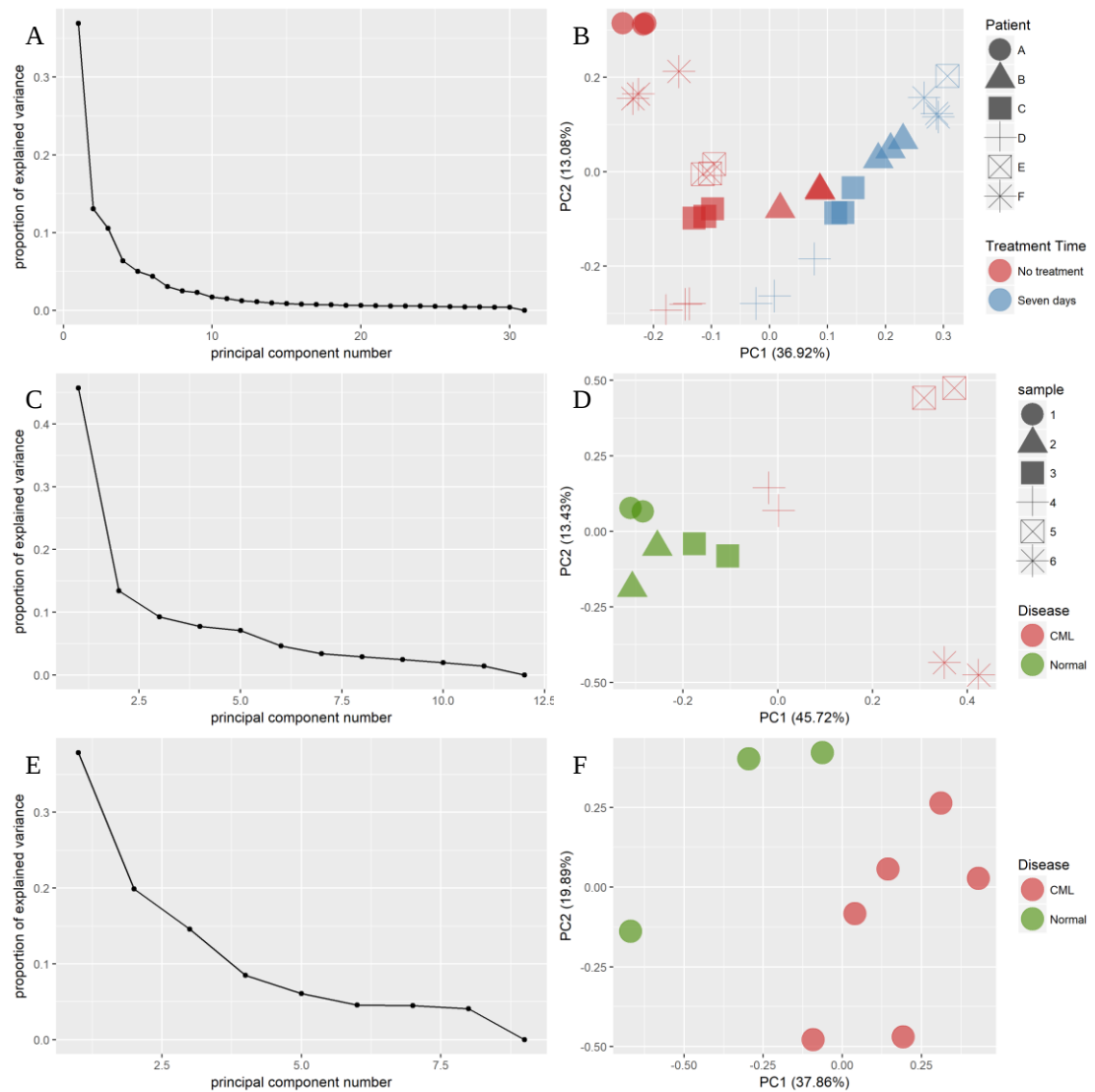


Figure 3-2. The different groups analysed in each dataset form distinct clusters in the PCA. (A) The scree plot of TKIFP shows a big difference in the explained variability between PC1 and PC2. (B) The samples in TKIFP cluster by patient and by treatment time, suggesting an effect of the treatment in the cells but also highlighting the need of controlling for the patient (to run a paired samples test). (C) The scree plot for CMLDV also shows a big difference in the explained variability between PC1 and PC2. (D) CML and normal form different clusters in CMLDV while the technical replicates (sample) cluster together. (E) In CMLMC the PC1 explained less variability than in the other datasets. (F) CML and normal HSCs form different clusters in CMLMC.

3.2.2 *Definition the threshold for no-change in microarray*

To find the genes that are not affected by TKI treatment, it was decided to use an equivalence test (CPMP, 2001). This kind of test is used in clinical studies (CHMP, 2010) but it is not commonly used in transcriptomic data analysis, which traditionally are focused in the differences between the studied groups. Equivalence tests assess the similarity of two different sets of measurements. A common approach in equivalence testing would be to calculate the difference between the mean of the $\log_2$ fold changes of the two groups for each gene/transcript and calculate the confidence interval 95 of the subtraction. For the two groups to be considered equivalent in the expression of a particular transcript/gene the confidence interval must cross zero and be positioned between two previously defined thresholds (one smaller and the other greater than zero; Figure 3-3). A key step in this methodology is the setting of the thresholds. Here, it was decided to use technical replicates to define them. Technical replication allows studying the variability of the data due to the handling of the samples and the limits of the technology as it provides different measurements for the same sample. This is very useful for setting the thresholds of an equivalence test as it provides a reasonable estimate of the variability produced by the technique and not the biology and therefore, the variability to be expected when there are no changes.

As all the datasets described above were generated using the same platform (Affymetrix HuGe 1.0 ST) it was decided to compare the technical replicates present in CMLDV and TKIFP to determine the null fold change distribution. As it was found that the different TKIs treatments (7d) of TKIFP dataset were more similar to each other (within the same patient) than the technical replicates of either TKIFP 0h or CMLDV (Figure 3-4A), it was decided to include them in the estimation of the technical variability. Assuming a p-value threshold of 0.05, we identified the 2.5th and 97.5th percentiles of the technical variation, which corresponded to the values -0.486 and 0.487 of the $\log_2$ fold-change. These values were rounded to one significant figure in the $\log_2$ scale: -0.5 and 0.5 (Figure 3-4B). All the values between these two percentiles represent $p \geq 0.05$, which is commonly accepted as a non-change value. Therefore, it was decided to use this interval, hereafter referred to as the “noise interval”, as the threshold for the equivalence test.

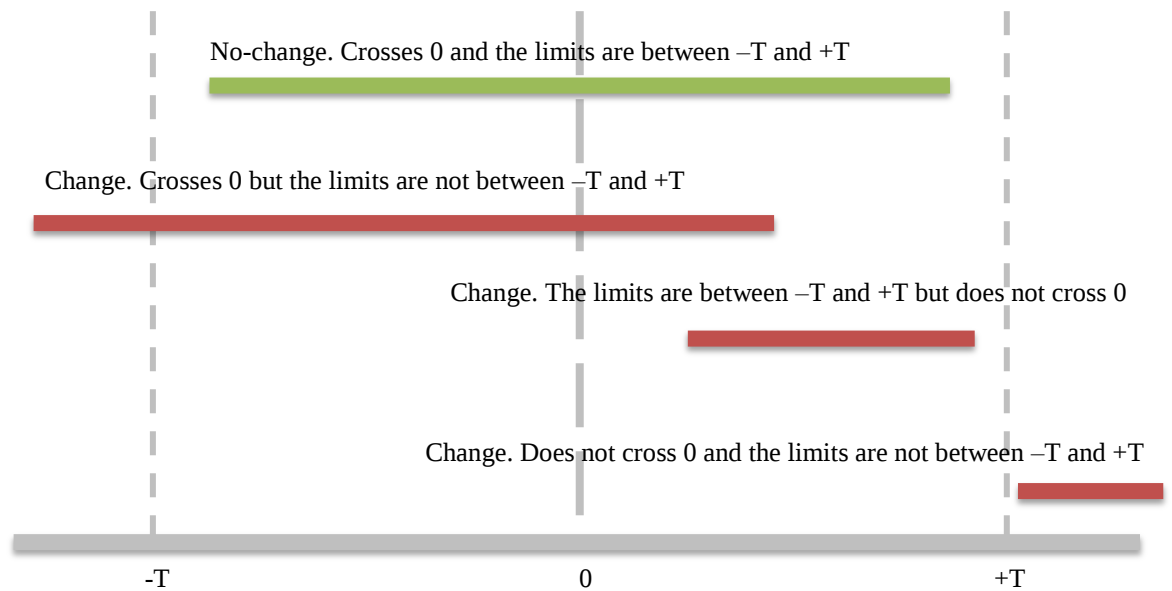


Figure 3-3. Explanation of what an equivalence test is. Equivalence tests are used when the alternative hypothesis (H_1) states that two values are equal. In order to inform about that, the confidence interval of the comparison is analysed as it must include zero and not a single value outside the interval delimited by the threshold (T) set by the researcher (the noise interval in the current study). An example of the confidence interval of a significantly non-changing comparison is plotted in green while three examples of comparisons where it is not possible to confirm equality are plotted in red.

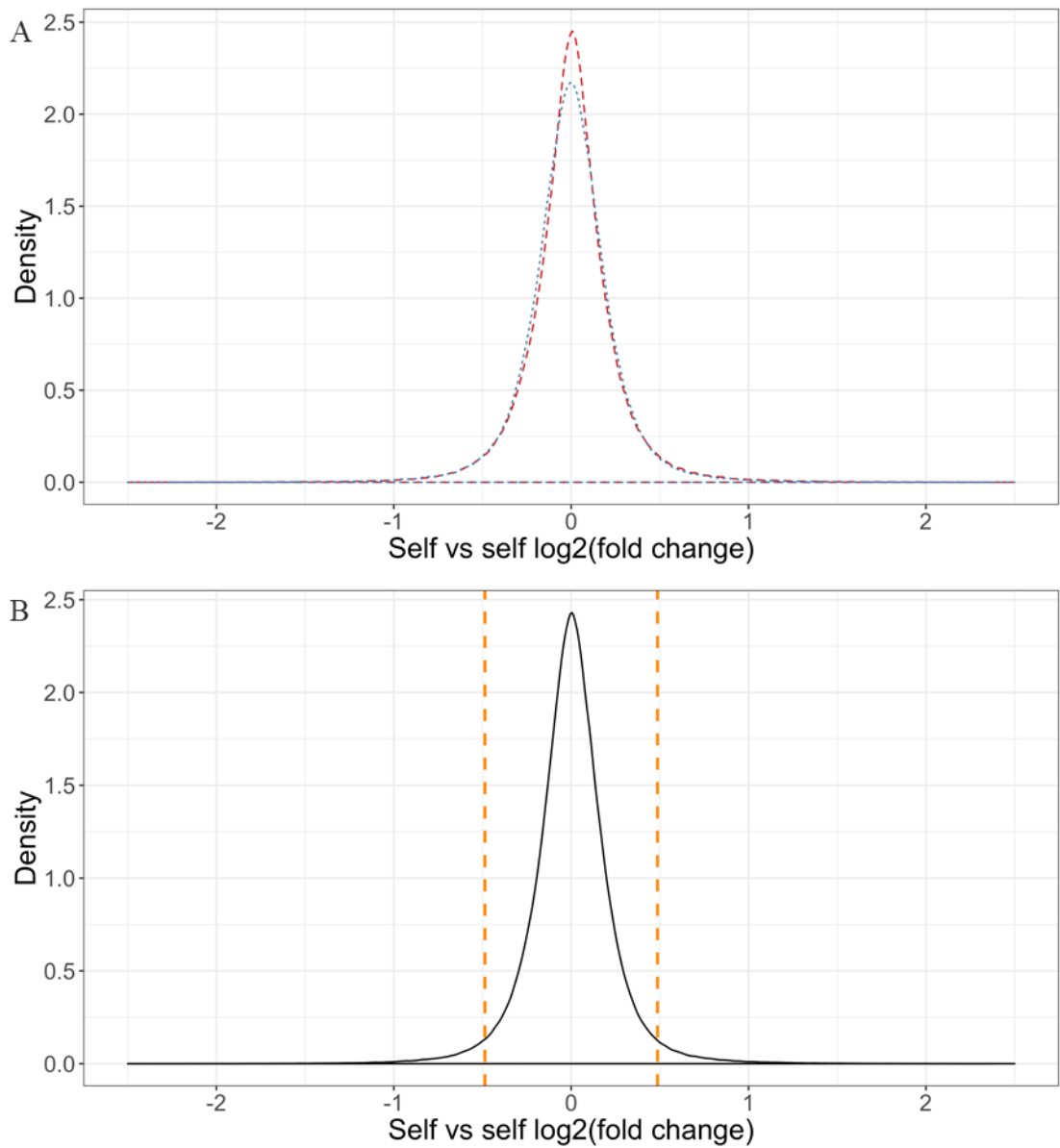


Figure 3-4. The noise interval was set between -0.5 and +0.5. (A) The distribution of the subtractions between the technical replicates in CMLDV (red) and the 7d TKIFP (blue) were similar. (B) Technical replicates were compared in order to build a distribution of the fold changes when comparing two samples that should be equal. The orange bars represent the percentiles 2.5 (-0.486 log₂ fold-change) and 97.5 (0.487 log₂ fold-change).

3.2.3 Differential gene expression and non-changing genes in the microarray datasets

limma (Smyth, 2004) was used for computing the $\log_2$ fold-changes in each of the three datasets and for computing the p-values of the differential expression analysis in CMLDV and CMLMC. TKIFP was analysed for detection of non-changing genes using an equivalence test, as described in the previous section (CPMP, 2001). Differential gene expression results were adjusted for multiple comparisons using BH method (Benjamini and Hochberg, 1995) and any gene with a q-value smaller than 0.1 was considered significant. The q-value was slightly relaxed from the usual 0.05 in order to increase the number of consistently changing genes (i.e. genes differentially expressed both in CMLDV and CMLMC). The requirement for the genes to be differentially expressed in both CMLDV and CMLMC reduced the concern about picking false positive genes (the product of the two 0.1 q-values is 0.01, still smaller than 0.05).

A total of 2,497 genes were differentially expressed in CMLDV and 888 in CMLMC (Figure 3-6A). To ensure accuracy and specificity in defining the *CML DE genes*, only those 527 genes differentially expressed in the same direction in both datasets were retained to represent the signature (Figure 3-6A, green shaded areas' overlap). On TKIFP, 5,706 genes were found significantly not-changing (Figure 3-6A, shaded blue). 60 genes were common to both lists (Figure 3-6A, circled in red); these 60 genes comprise the initial candidate list representing TKI₆₀ signature (Figure 3-6B).

Further analysis of the list using CPDB (Kamburov et al., 2013) revealed an overrepresentation of genes involved in olfactory transduction (*ANO2*, *OR2L5*, *OR2L2*, *OR2L3*, *OR2AK2* and *OR2L8*; $q=0.013$) and cell adhesion molecules pathways (*SELL*, *CDH2* and *ESAM*; $q=0.028$). Olfactory receptor (OR) genes were all downregulated but *ANO2*, a calcium-activated chloride channel. *ANO2* is located in chromosome 12p just downstream of *VWF*, a haemostasis factor, and their expression as shown to be related in other conditions (Schneppenheim et al., 2007).

All the OR genes in the TKI₆₀ signature clustered together in the same region of the genome. Previous reports suggest that the expression of OR genes is affected by neighbouring genes (Feldmesser et al., 2006). Further investigation of this region of the genome using the genome browser of UCSC (Kent et al., 2002) revealed *TRIM58* as a neighbouring gene of the TKI₆₀ OR genes (Figure 3-5). Additionally, *TRIM58* promoter's histones are highly acetylated in K562 cells (ENCODE Project Consortium, 2012) and it has been found to be an important regulator of erythropoiesis (Thom et al., 2014). With

this in mind, it was decided to substitute the OR genes for *TRIM58* in the downstream analysis even when it was not present in the initial list of 60 genes.

ESAM has been found to be an important marker for HSCs both in human (Ishibashi et al., 2016) and mouse (Yokota et al., 2009). It has also shown to have higher expression in high proliferating but repopulating HSCs (Sudo et al., 2012) and it has been reported as a potential AML LSC marker after a transcriptomic and a proteomic screen (Bonardi et al., 2013). Similarly, L-selectin (*SELL*) and N-cadherin (*CDH2*) seem to be involved in HSC maintenance and regulate differentiation (Zhi et al., 2016, Agnihotri et al., 2017).

The finding of *MIR10A* downregulated in the list of TKI₆₀ genes increased the confidence in the list as its TKI independent downregulation in CML has previously been reported by an independent group (Agirre et al., 2008). The decrease of *MIR10A* expression is linked to an increase in the protein levels of *USF2* and an increase in cell proliferation.

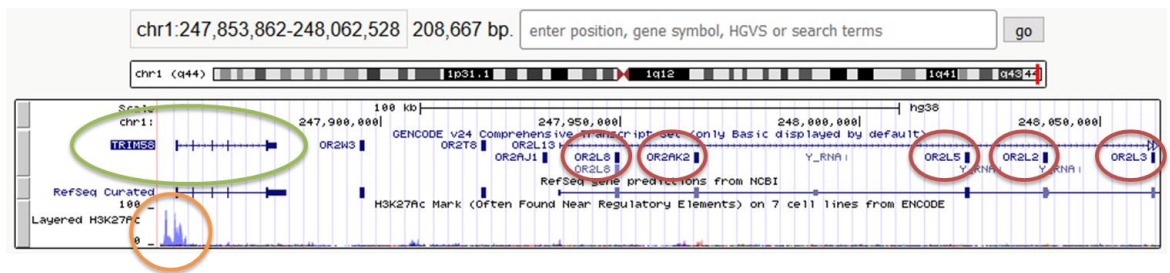


Figure 3-5. The 4 olfactory receptor (OR) genes present in the TKI₆₀ list cluster together in the genome at 1q44 (red circles). The presence of so many OR genes lead to a detailed screening of their chromosomal neighborhood, where *TRIM58* (green circle) was found. *TRIM58* presented high H3K27 acetylation (orange circle) in K562 cells, suggesting a high expression level in this cell line.

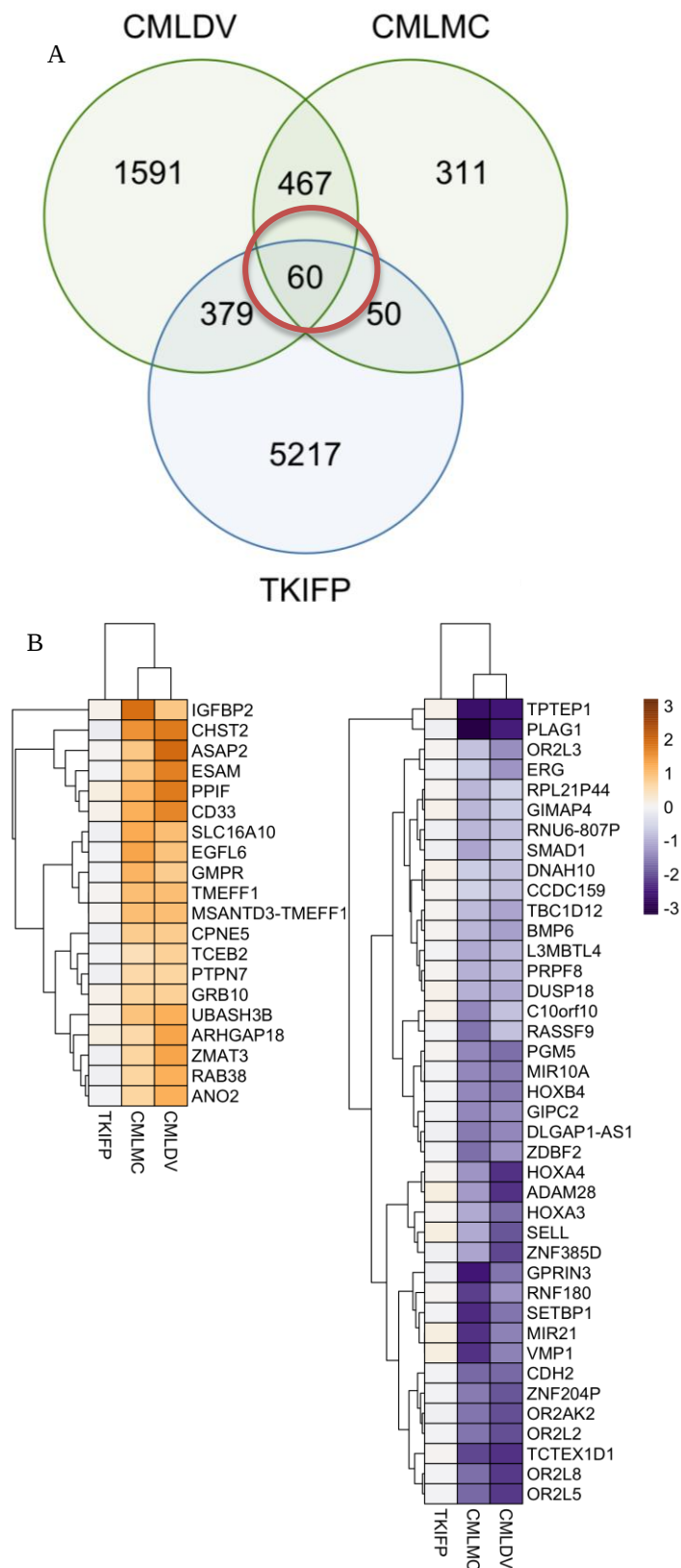


Figure 3-6. A total of 60 genes were identified as TKIi. (A) After the analyses of the 3 microarray datasets, 60 genes were found to be present in the 3 lists, that is, to be TKI₆₀. (B) Heatmaps showing the log₂ of the fold changes of the comparisons for the 60 genes that are present in all 3 lists (genes upregulated in CML are shown on the left, genes downregulated in CML are shown on the right).

3.3 Validation of the TKli signature

3.3.1 *In silico validation of the TKli signature*

The probability of obtaining a 60 gene signature was tested using both Monte-Carlo and hypergeometric distribution. Monte-Carlo is based in the repetition of the analysis multiple times but randomizing or permuting the allocation of samples into the groups (i.e. the classification into normal or CML for CMLDV and CMLMC datasets) (Metropolis and Ulam, 1949). More information in the Methods (2.2.3.2). The number of TKli probe sets instead of genes was used in order to accelerate the experiment (annotation with *biomaRt* requires connection with Ensembl). After performing 10,000 permutations of the sample names of both CMLDV and CMLMC, 12 reached the number of probe set IDs (58) in the actual TKli signature, giving a significance of $p=0.0012$ (Figure 3-7C). Not a single permutation returned a number of overlapping probe set IDs higher than 58. As further validation, the probe set IDs in all three datasets were permuted (i.e. the probe set IDs were assigned to a different row of expression values) and overlap between the three datasets was assessed. After 100,000 permutations, the maximum number of overlapping probe set IDs was 28, way below the number of TKli probe sets (58; Figure 3-7D).

The hypergeometric distribution returns the probability of obtaining a certain number of cases of one condition from a particular set of cases where two conditions are mixed (Methods 2.2.3.2). Here, the hypergeometric distribution was used to calculate the probability of getting the same or greater/lower number of cases in an overlap between two sets of significant genes: (I) genes differentially expressed in both CMLDV and in CMLMC and (II) the 527 common deregulated genes in CML and the genes that do not change in TKIFP. It was found that the probability of getting the same number of genes (or more) overlapping between CMLDV and CMLMC was close to zero (Figure 3-7A) while the probability of getting an overlap of 60 genes (or less) between TKIFP and the 527 genes common in both CML datasets was also close to zero (Figure 3-7B).

Overall, these results suggest that the identification of 60 genes (or 58 probe set IDs) comprising the TKli₆₀ signature is unlikely to have been obtained by chance. The Monte-Carlo analysis calculates the probability of getting the same number of genes or larger by chance using the data and methodology than in the original analysis, which is very low. Furthermore, the hypergeometric distribution supports the hypothesis as the number of common genes in CMLDV and CMLMC was higher than expected by chance, which correlates with the fact that in both datasets the same CML DE genes comparison was performed (using the same methodology). However, the number of genes that overlap

between the 527 CML genes and the significant genes in TKIFP is lower than expected by chance. It is important to highlight that although the number TKIi genes found in the analysis is lower than the expected by chance, the biological context should be considered here. It is known that CML cells, and in particular committed and progenitor cells, are sensitive to TKI treatment (Corbin et al., 2013, Holtz et al., 2002). It has been suggested that TKI treatment restores the normal transcriptome in CML cells by inhibiting the BCR-ABL1 TK activity (Hamilton et al., 2012), restoring normal haematopoiesis (Holtz et al., 2002). As the 527 CML DE genes are potentially affected by the TK activity of BCR-ABL1, inhibition of this TKI activity should restore most of these genes' normal expression levels. This correlates with the significant reduction in the number of TKIi genes in the second hypergeometric distribution. Thus, the TKIi genes are actually an exception to the rule, just as the BCR-ABL1 TK independent genes are expected to be.

3.3.2 In vitro validation of TKIi signature: primer optimization

In order to be more confident in the reproducibility of the gene expression signature, the results were further validated using an independent cohort of patients (Table 2-1, Table 2-2) and their gene expression was assessed by qPCR.

PCR primers do not always have the same efficiency. Although theoretically PCR reactions should duplicate the number of molecules of the target sequence of DNA or cDNA (amplicon), this is not always the case (Karlen et al., 2007). This is a problem when the expression of different genes is compared by qPCR, as different primers can amplify their target sequence to different extents in the same cycle of the reaction. To confirm that primer efficiencies are consistent it is necessary to test primer efficiency using a series of cDNA dilutions of known concentration. Doing that, it is possible to determine the efficiency of the reaction in the range of concentrations tested and select those primers with efficiencies close to 100% (that is, that duplicate the amplicon in every cycle). Determination of efficiency was performed using Equation 3 using standards for 50, 25, 10, 2.5 and 0.5ng of cDNA prepared from RNA extracted from wild type (WT) K562 cells in a 10 μ L reaction.

Primers were designed for protein coding and miRNA genes but not for pseudogenes or other non-protein coding genes. Primers for a total number of 46 genes were tested. From those, primer pairs for 35 genes presented efficiencies in the range of 80 to 120% in an interval of at least 3 concentrations containing 10ng per reaction (1ng/ μ L of reaction) (Figure 3-8).

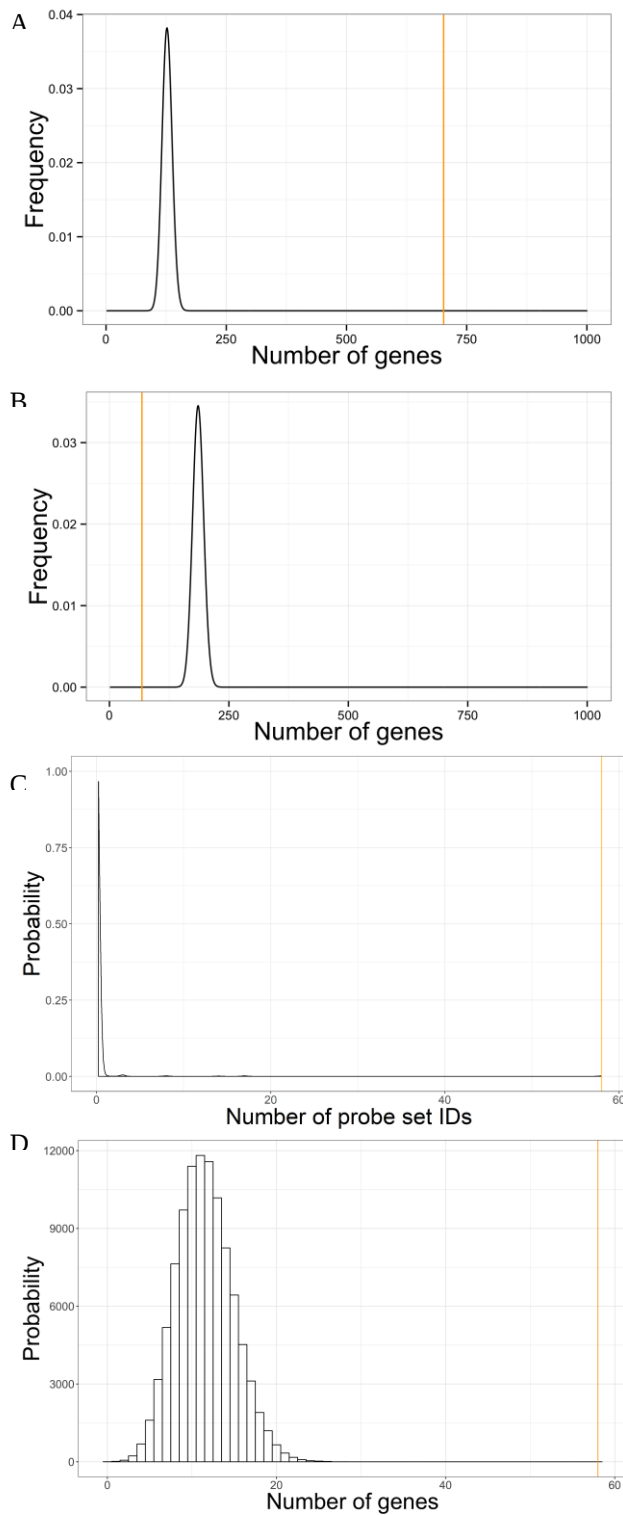


Figure 3-7. The TKI₆₀ signature is unlike to be generated by chance. The orange bar always represents the number of genes/probe set IDs in the actual data. (A) Distribution of the probability of getting 527 common genes in the two CML datasets by chance using hypergeometric distribution. (B) Distribution of the probability of getting 60 genes that do not respond to TKI treatment in the 527 CML genes using hypergeometric distribution. (C) Distribution of the number of TKI₆₀ probe set IDs obtained after performing 10,000 permutations on the samples names of CMLDV and CMLMC at the same time and calculating the overlap of the differentially expressed genes in the permuted datasets and the non-modified TKIFP (Monte-Carlo). (D) Distribution of the number of TKI₆₀ genes obtained after performing 100,000 permutations of the gene names in all the three datasets.

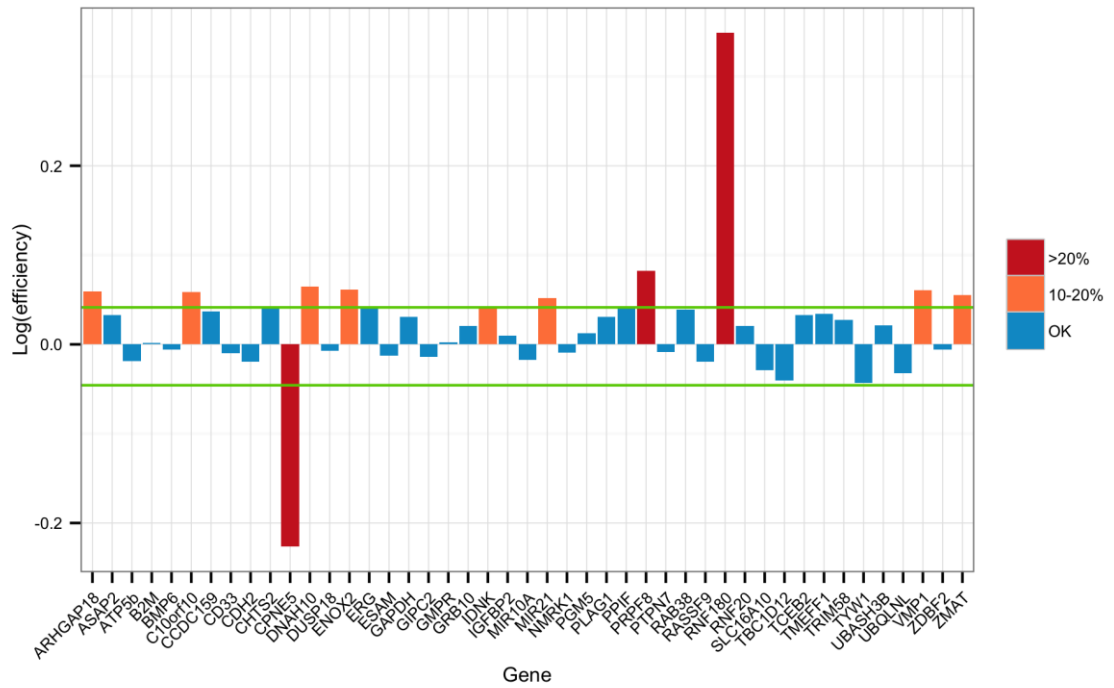


Figure 3-8. Primer efficiencies in K562 cells. The green lines represent the interval between 90 and 110% efficiency, which is considered the optimal efficiency for a pair of primers. However, the primers with efficiencies 80-120% were also included in later experiments.

$$\text{Amplification efficiency} = -1 + 10^{1/\text{slope}}$$

Equation 3. Primer efficiencies equation. Using the slope obtained after plotting the Ct value (y axis) against the base 10 logarithm of the concentration of DNA present in the reaction (x axis) at different concentrations it is possible to determine the efficiency of the primers under the conditions used in the reaction.

3.3.3 Validation of the TKli signature by qPCR in CD34⁺ cells

Validation of the TKli₆₀ signature was performed using a cohort of independent patient samples from those used in the microarray experiments from CML (n=5) and nCML (i.e. mantle cell lymphoma and lymphoma, haematological malignancies that do not affect the HSC population; n=2) patients as well as from normal donors (i.e. allogenic bone marrow donors, treated with G-CSF; n=2). All the samples were enriched for CD34⁺ cells as described in the Methods section (2.2.1.3). CML samples were treated with 5μM IM for 7 days culturing them in serum free media in the absence of growth factors. Imatinib was reapplied after 96 hours without changing the media and sorted for viable cells (DAPI⁺) on

day 7. RNA was extracted at baseline (just after CD34⁺ enrichment) and after sorting, for both the IM treatment and no drug control arms.

Gene expression of the 39 genes was measured using a 48.48 Fluidigm chip and *ENOX2*, *GAPDH*, *RNF20* and *TYW1* as reference genes. A new noise interval specific to the Fluidigm data was calculated using the technical replicates included in the chip ($Ct_{\text{replicate1}} - Ct_{\text{replicate2}}$). As the reduced number of comparisons did not allow calculating a symmetrical noise interval, all the Ct difference values were transformed into their absolute value. This should not have an effect on the results as the technical replicates could be arbitrarily exchanged in the subtraction and it would still have the same meaning (difference between two technical replicates). Thus, the percentile 95 was calculated (instead of the percentiles 2.5 and 97.5). The noise interval was found to be from -0.49 to 0.49 (rounded to -0.5 to 0.5) differences in Ct, similar to the one previously described for the microarrays (Figure 3-9). Using this interval, 12 genes – *PPIF*, *TRIM58*, *CD33*, *CHST2*, *PRPF8*, *ASAP2*, *GIPC2*, *UBASH3B*, *TCEB2*, *GRB10*, *ERG* and *MIR10A* – were found to be non-changing in CML CD34⁺ cells after IM treatment (Figure 3-10A).

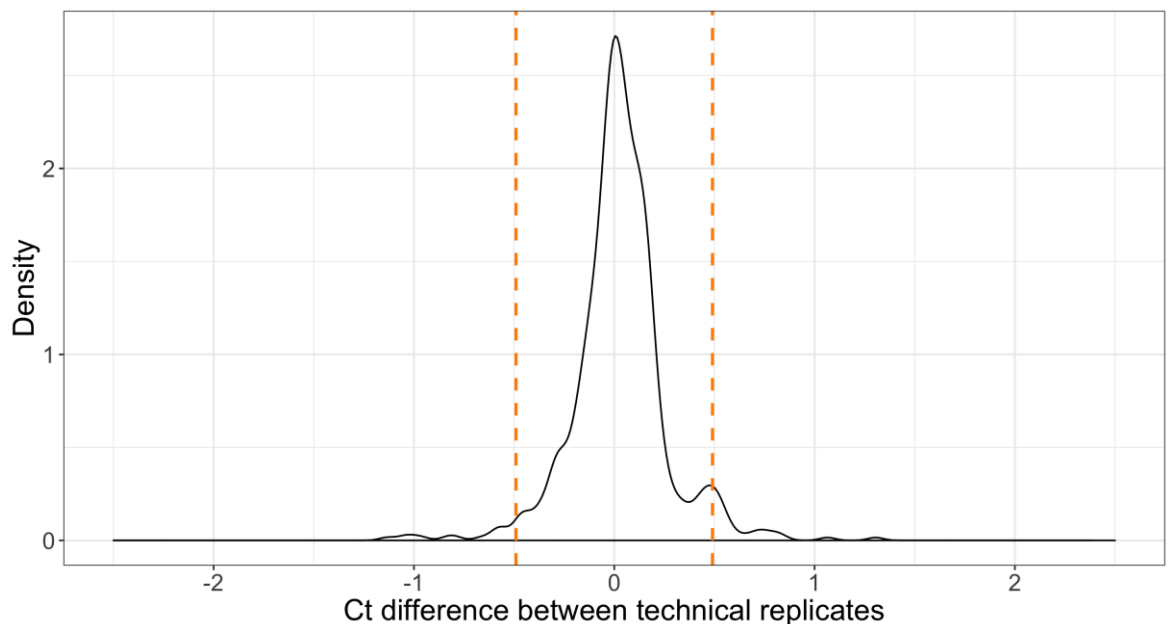


Figure 3-9. The noise interval in the 48.48 Fluidigm chip was similar than in the microarrays. Technical replicates were compared in order to build a distribution of the fold changes when comparing two samples that should be equal. The orange bars represent the mean of the absolute values of the percentiles 2.5 and 97.5 (-0.49 and 0.49 difference of Ct).

Originally, the analysis was performed using both the nCML and the normal samples as controls to compare against CML samples. The pool of nCML and normal is referred to as ‘nCMLp’ for the rest of this section. Performing this analysis, 18 genes (Figure 3-10B) were differentially expressed in CML compared with nCMLp, of which 6 were not affected by IM treatment (Figure 3-10C). However, *UBASH3B* was found upregulated in the microarray analysis while it was found downregulated after the qPCR analysis, leaving a final list of 5 genes: *MIR10A*, *GIPC2*, *TRIM58*, *ERG* and *CD33*.

However, a closer look at the differences between normal donor samples and nCML samples revealed a trend showing important gene expression differences between the two groups (Figure 3-11A). This suggests that the analysis should be repeated using only truly normal samples as controls (the two allogenic bone marrow donors). After redefining the control samples and repeating the analysis as previously described, 13 genes were found to be differentially expressed in CML compared with normal (Figure 3-11B), of which 4 were not affected by IM treatment: *ERG*, *CHST2*, *PPIF* and *CD33* (Figure 3-11C, Table 3-1). Although the small number of normal samples reduces the power of the experiment the differences observed between the normal donors and the nCML suggested to only use the normal donors as controls.

Table 3-1. Gene expression changes of the TKI_{i60} genes with successful amplification in Fluidigm. Gene expression changes are shown in log₂ fold-change. The genes validated by qPCR are highlighted in red. CML qPCR column shows the mean gene expression changes between CML and normal samples. TKI qPCR column shows the mean gene expression changes between NDC and IM 5μM. CMLDV, CMLMC and TKIFP are the respective microarray datasets. *RNF180* did not present detectable expression for both the NDC and the IM 5μM for any single sample and it is marked as NA for the TKI qPCR column.

Gene	CML qPCR	TKI qPCR	CMLDV	CMLMC	TKIFP
ARHGAP18	0.94	-0.50	1.32	0.61	0.13
ASAP2	-2.14	0.03	1.94	0.85	0.04
BMP6	-6.47	0.71	-1.18	-0.93	0.04
C10orf10	-4.27	1.86	-0.82	-1.42	0.09
CCDC159	-0.72	1.07	-0.78	-0.58	0.03
CD33	2.43	-0.11	1.61	1.16	0.10
CDH2	-6.91	-0.99	-1.79	-1.77	-0.02
CHST2	2.21	-0.08	1.75	1.48	-0.13
DNAH10	-4.76	1.42	-0.79	-0.67	0.07
DUSP18	-4.75	0.55	-1.04	-1.00	0.09
ERG	-1.65	0.44	-1.30	-0.66	-0.06
ESAM	-1.02	-2.37	1.68	0.95	-0.04
GIPC2	-3.65	0.07	-1.37	-1.42	-0.02
GMPR	2.09	-1.48	0.82	1.12	-0.01
GRB10	0.40	0.27	0.71	0.65	0.07
IGFBP2	-0.06	-1.35	0.88	1.91	0.10
MIR10A	-4.24	0.48	-1.56	-1.45	-0.06
MIR21	-1.24	2.60	-1.52	-2.29	0.16
PGM5	-5.40	2.10	-1.69	-1.46	0.01
PLAG1	-5.73	2.09	-2.48	-3.07	-0.10
PPIF	1.21	-0.31	1.78	1.10	0.14
PRPF8	0.00	-0.05	-0.93	-0.98	0.05
PTPN7	1.40	0.82	0.70	0.62	-0.07
RAB38	0.70	-1.07	1.18	0.69	-0.12
RASSF9	-4.85	1.35	-0.83	-1.63	-0.02
RNF180	-4.21	NA	-1.33	-2.17	0.01
SLC16A10	1.85	-1.45	1.01	1.22	-0.07
TCEB2	0.50	0.13	0.73	0.47	-0.05
TMEFF1	0.15	-1.78	0.97	1.01	0.01
UBASH3B	-2.15	0.08	1.16	0.93	0.13
VMP1	-0.48	0.73	-1.52	-2.29	0.16
ZMAT3	0.44	0.66	1.33	0.70	-0.07

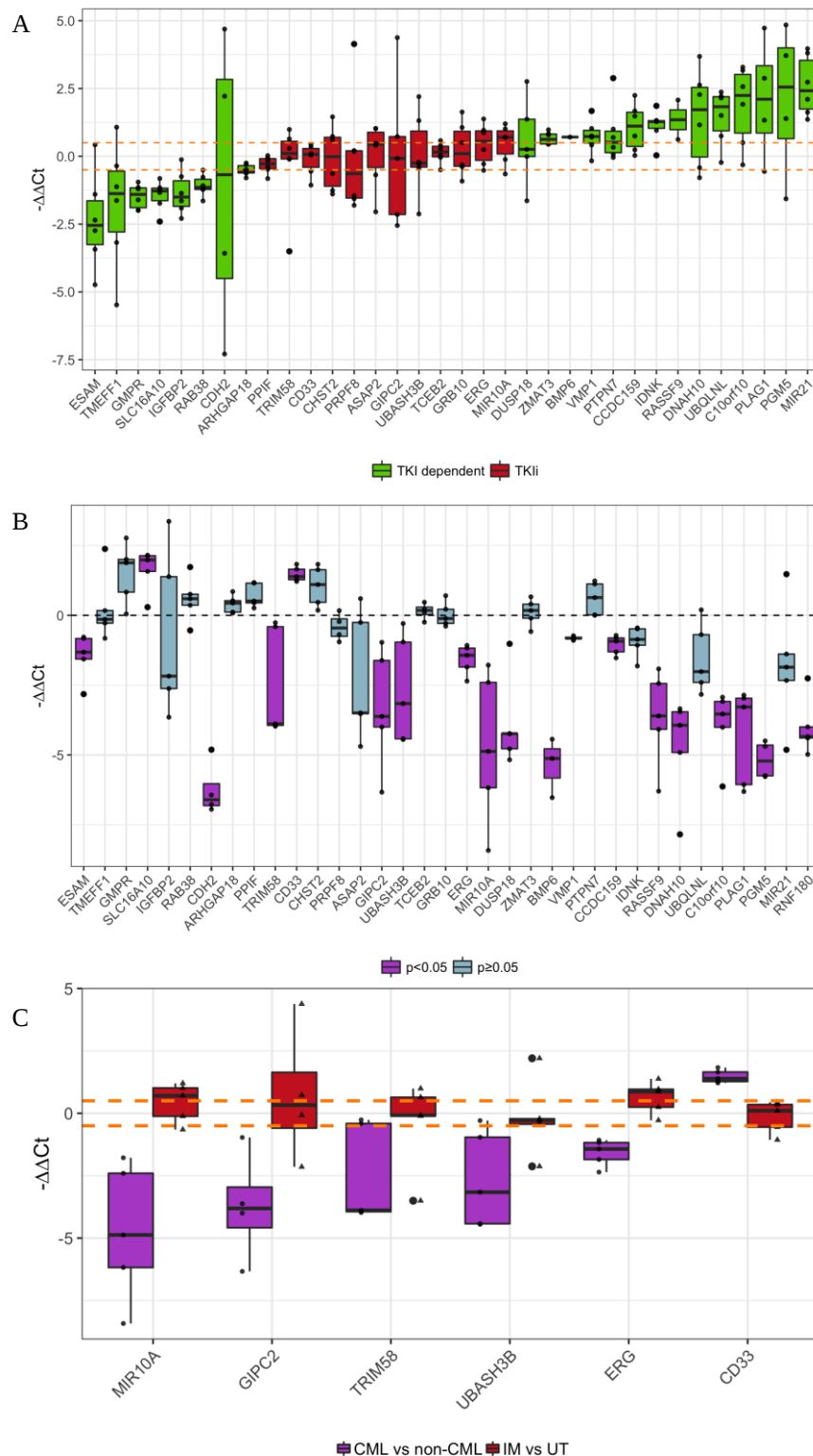


Figure 3-10. A list of 6 genes was found to be TKIi using nCMLp samples as control. (A) Boxplots showing the gene expression differences of the 39 genes tested using Fluidigm following 7 days of culture in SFM without growth factors in the presence and absence of 5 μ M IM. The horizontal, orange dashed lines define the noise interval. The genes with mean $-\Delta\Delta Ct$ occurring within the noise interval are highlighted in red. (B) Gene expression differences between CML and nCMLp in all the genes analysed by qPCR. The boxplots coloured in purple represent those genes that are significantly differentially expressed. (C) Those genes that are both differentially expressed in CML compared with nCMLp and not affected by the 7 days IM treatment. *UBASH3B* is downregulated in the Fluidigm chip while it was upregulated in the microarray analysis, so it was discarded.

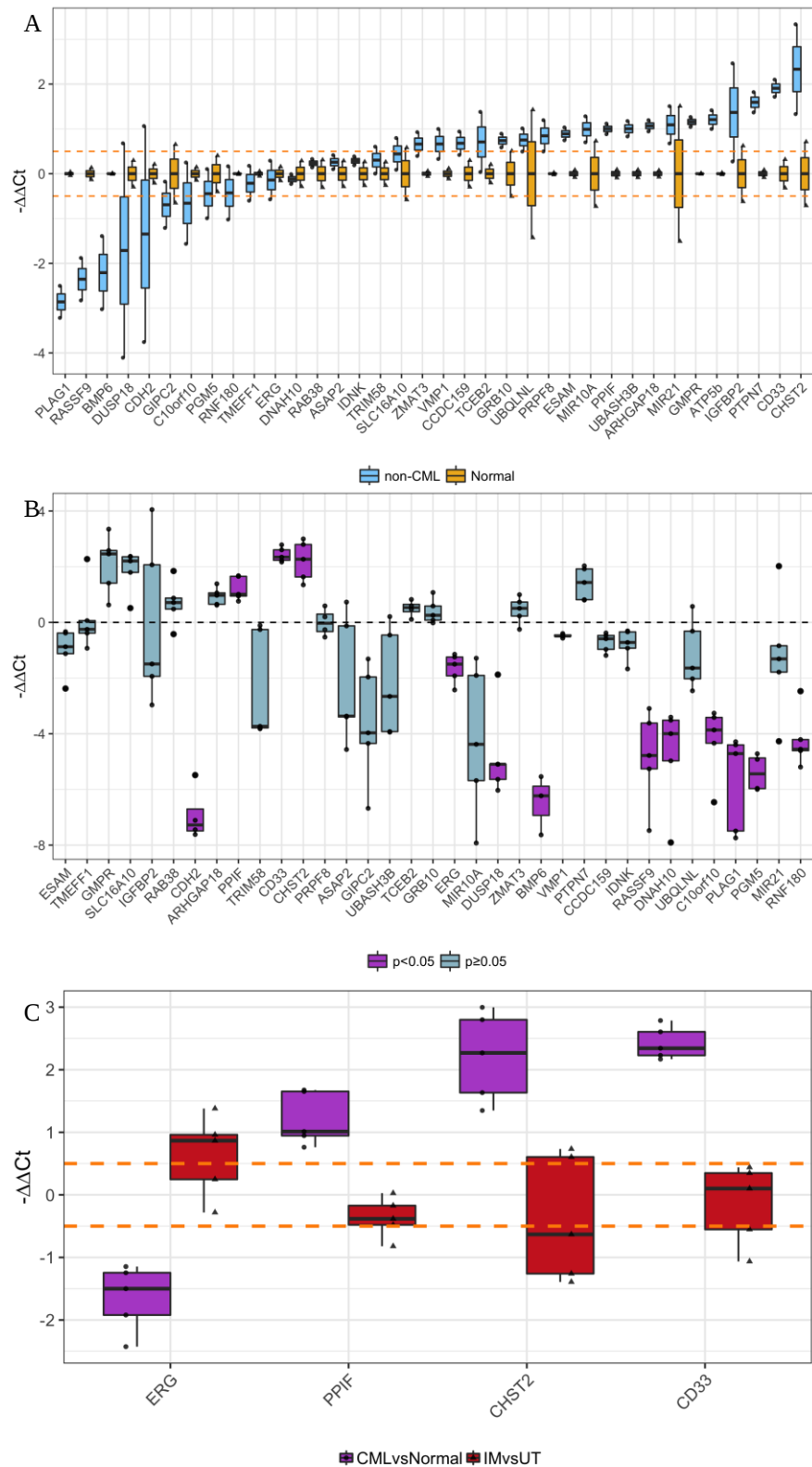


Figure 3-11. A list of 4 genes was found to be TKIi in using normal samples as control. (A) Differences between normal donor samples and those from nCML malignancies. Most genes were not equal in both groups as they were not confined within the noise interval (orange lines). $-\Delta\Delta Ct$ was calculated using the mean of the normal donors as reference. (B) Gene expression differences in all the genes analysed by qPCR between CML and the normal donors. The boxplots coloured purple represent the genes that were significantly differentially expressed. (C) Shortlist of the genes that were both differentially expressed in CML compared with normal (purple) and not affected by the 7 days IM treatment (red). The data for each of these comparisons are shown next to each other for each gene.

3.4 Discussion

The high efficacy of TKIs for controlling CML but their lack of success at eradicating the LSCs (Mahon et al., 2010, Ross et al., 2013, Clark et al., 2017) requires new therapeutic approaches that are able to eradicate CML LSCs. The success of TKIs means that any new therapy in CML should be designed to cooperate with, not substitute for, TKIs as withholding such an effective medicine would not be ethical and clinical trials would have difficulty in recruiting patients. Also, the limited toxicity of TKIs on non-leukaemic HSCs means that any novel therapy must have a low toxicity profile in order to maintain the quality of life of the patients.

Previous work has been done targeting members of other signaling pathways important for LSCs survival, such as PP2A (Neviani et al., 2013), EZH2 (Scott et al., 2016), p53 and c-MYC (Abraham et al., 2016), showing promising results. However, the discovery of these new targets was performed on untreated cells, which might lead to elucidation of target pathways that may be normalized by TKI. Thus, some of the benefits achieved by each of the drugs might be redundant with those of TKI and a similar or improved effect could be achieved with reduced toxicity targeting pathways not affected by TKI treatment. The novelty of the work presented in this chapter lies in the discovery of a new gene signature (TKI_{i60}/TKI_{i4}) whose members' expression is unaffected by TKI treatment but is disease specific (differentially expressed in CML compared with normal controls).

The new approach described in this chapter may lead to more effective screening for potential drug targets that would cooperate with TKI treatment in CML LSCs. Here, a list of 4 genes (*ERG*, *CHST2*, *PPIF* and *CD33*) is presented as potentially valid drug targets in CML LSCs uncovered after the analysis of microarray datasets and validation using microfluidics qPCR.

There are no well-established protocols for detecting significantly equivalent levels in most disciplines and borrowing methodologies from other disciplines requires adaptation. In this chapter a new protocol for detection of non-changing genes is presented by adapting the equivalence test previously used in clinical trials (CPMP, 2001) to microarray analysis. By calculating the variability between pairs of technical replicates it was possible to infer the normal variability expected by the instrument. This allowed setting a non-arbitrary threshold for the equivalence test and provides enough flexibility to be used among different platforms (e.g. microarray and microfluidics qPCR). The *in silico* validation of the number of TKI_{i60} genes also showed that these results are unlikely to be obtained by chance, further supporting the designed protocol.

The relative low validation rate from the microarray analysis in the qPCR (4/35) could be explained by different factors. The first is the different cell type analysed: while the CMLMC and CMLDV datasets included primitive LSCs ($CD34^+CD38^-/CD34^+CD38^-CD90^+$) the experiments described here were performed using $CD34^+$ cells, which contain primitive populations but primarily comprise more committed cells. The main reason to work with a different population was the availability of material. While $CD34^+$ cells are already a rare population of cells (about 0.02% of leukocytes) (Kikuchi-Taura et al., 2006), $CD34^+CD38^-/CD34^+CD38^-CD90^+$ populations are even rarer (about 5% of the $CD34^+$ population) and working with them would require large patient samples, especially when treating with cytotoxic drugs. Secondly, the original microarray dataset included three different time points (0h, 8h and 7d) but only the 0h time point was treatment naïve with no valid controls at 8h or at 7d. This meant that the time point was confounded with the treatment (NDC/treated) with no means of controlling for it. For the validation a NDC was included at 7d in order to control for the effects that the time point may have on the cells, such as cultural artefacts and differentiation. This, while improving the overall experimental design, also introduced a difference between the microarray dataset and the validation data.

Another point to consider is the use of only healthy $CD34^+$ allogeneic donors as controls. The availability of this kind of sample for research is limited and, therefore, other samples are usually selected for this purpose. Generally, haematological malignancies such as lymphoma and myeloma (nCML) do not have their $CD34^+$ population affected by the disease. That means that the $CD34^+$ population is similar to that of a healthy person although these patients might have been on treatment already, which can have an effect in gene expression. However, no studies have demonstrated this similarity and when the gene expression levels of the 35 genes analysed in the microfluidics qPCR for the nCML samples and the healthy donors were compared, most genes did not fall within the limits for non-change. This led to the decision of using only healthy donors as controls for the experiment despite of the loss of statistical power (the experiment was designed with 4 controls but 2 were dropped on this decision).

The final list of TKI₄ genes comprises *ERG*, *PPIF*, *CHST2* and *CD33*. *ERG* is a transcription factor that has already been found to participate in other cancers such as AML (Martens, 2011, Knudsen et al., 2015), pre-B-ALL (Clappier et al., 2014) and prostate cancer (Adamo and Ladomery, 2016), where its activity is very important for the development of the tumour. However, it seems to be downregulated in most cancers through hypermethylation of the two CpG islands in its promoters (Adamo and Ladomery,

2016). ERG interacts directly with GATA2 and RUNX1 (Wilson et al., 2010) and promotes stem cell maintenance and quiescence in HSCs via repression of MYC (Knudsen et al., 2015). ERG belongs to the ETS family of transcription factors and binds to a motif between 15 and 20 base pairs, becoming more promiscuous after suffering post-translational modifications or heterodimerisation (Adamo and Ladomery, 2016). Its three different promoters allow for a transcriptional diversity, reaching up to 30 different transcripts and 15 proteins, which regulate different targets. ERG cooperates with histone deacetylases (HDACs) and the polycomb complex recruiting EZH2 and inhibits the CBP/p53 pathway (Adamo and Ladomery, 2016).

CHST2 is an essential component of the cell sulphonation machinery (Kawashima et al., 2005, Uchimura et al., 2005). Sulphonation is a post-translational modification mechanism on glycosylated proteins. This process is fundamental for the synthesis of valid L-selectin ligands – which includes CD34 – and, therefore, it is necessary for the homing of lymphocytes in the lymph nodes (Kawashima et al., 2005, Uchimura et al., 2005). Additionally, *CHST2* has been found to be upregulated in low risk B precursor ALL with *ERG* intragenic deletions, in contrast with the high risk BCR-ABL1⁺ patients (Harvey et al., 2010). This could suggest a role for cell-cell interactions in CML LSC persistence.

PPIF is a key initiator of autophagy in the mitochondria as it participates in the mitochondria transition pore. *PPIF*^{-/-} cardiomyocytes are unable to initiate autophagy, while PPIF overexpressing cells increase autophagy even under non-starvation conditions (Carreira et al., 2010). During ischemia, PPIF forms a complex with p53 that activates necrosis and it has been shown that *PPIF*^{-/-} mice exhibit protection against ischemic brain disease (Vaseva et al., 2012). Conversely, a co-immunoprecipitation study showed that PPIF associates with BCL2 (Eliseev et al., 2009). This association protects the cell from apoptosis, probably by the regulation of cytochrome c release (Eliseev et al., 2009). Thus, PPIF participates in the regulation of cell death by both protecting cells from apoptosis and by increasing necrosis. The increase in the levels of PPIF and its pro-necrotic activity may be countered by the reduced activity of p53 pathway in CML LSCs (Abraham et al., 2016). Therefore, high levels of PPIF in CML LSCs may only protect against apoptosis without promoting necrosis. Cyclosporine A, a widespread immunosuppressant drug, targets PPIF and it has been shown to inhibit its activity (Eliseev et al., 2009, Vaseva et al., 2012).

CD33 is a myeloid cell surface marker that has been exploited as a therapeutic target in AML because of its high expression in AML blasts. Between 2000 and 2010 gemtuzumab-ozogamicin (GO), commercialised as Mylotarg by Pfizer, was used routinely for treating

AML, but was withdrawn from the market after a clinical trial showed hepatotoxicity in a number of patients (Jurcic, 2012). However, other clinical trials have demonstrated that GO is highly beneficial for some subsets of patients, especially those with low risk cytogenetics (Jurcic, 2012) and recently it has been re-approved by the US FDA for its commercialisation (Jen et al., 2018a). Although normal HSCs can present expression of CD33, it is believed that CD33 is not expressed in normal HSCs in the presence of LSCs, as no effect in normal haematopoiesis has been detected in patients treated short-term with GO (Pearce et al., 2006). However, when patients are treated long-term with GO some of them develop thrombocytopenia, probably because of the re-expression of CD33 in normal stem or progenitor cells (Pearce et al., 2006). A previous publication already reported the high expression of CD33 in CML LSCs (Herrmann et al., 2012).

In summary, the TKI₄ gene signature described in this chapter comprises *ERG*, *PPIF*, *CHST2* and *CD33*. Both CD33 and PPIF can be targeted using commercially available drugs that are already approved for use in the clinical practice. The effect of gemtuzumab-ozogamicin (GO) on CML CD34⁺ cells targeting CD33 will be discussed on Chapter 5 of this thesis. Additionally, Chapter 4 discusses the role of the TKI₄ signature as a biomarker for TKI response and prognosis.

4 Results (II): Investigation of the potential role of the TKI independent signature as a biomarker in CML LSCs

4.1 Introduction

The uncovering of the TKIi signature in the previous chapter reinforced the hypothesis that CML LSCs possess de-regulated genes that are independent of the TK activity of BCR-ABL1. However, its existence does not necessarily mean that it has any real application in a clinical environment. The aims of this chapter is to (I) assess the TKIi signature as a potential biomarker for TKI response, disease aggressiveness and phase of the disease and (II) investigate if the differential expression of the TKI signature in CML CP is maintained over accelerated phase (AP) and blast crisis (BC) using already existing microarray datasets.

Previous publications have already investigated the existence of gene expression signatures as biomarkers for TKI response (McWeeney et al., 2010), disease aggressiveness (Yong et al., 2006) and phase of the disease (Cramer-Morales et al., 2013) (CMLMC). In this chapter, these data will be exploited to evaluate the value of the TKIi signature *in silico*. The different datasets were processed for finding gene expression differences between the contrasts of interest (TKI response, disease aggressiveness and phase of the disease) and the results compared with the TKIi signature mentioned in the previous chapter.

The different microarray platforms used for each dataset make direct comparison between the two datasets introduced in this chapter (Yong et al., 2006, McWeeney et al., 2010) and the TKIi₆₀ list more complicated. Summarising the values of the probes representing the same gene, as was done in the previous chapter, provides a more stable value for the expression of each gene. However, each microarray platform contains a different set of probes that do not necessarily bind to the same part of the transcript. Therefore, comparing summarised gene values from different platforms may provide results from a different set of transcripts (splicing variants). In order to solve this issue the comparisons between different microarray platforms were performed using RefSeq transcripts (O'Leary et al., 2016) instead of genes.

To assess the value of the TKIi signature as a biomarker for TKI response, disease aggressiveness and disease progression, the differentially expressed transcripts in the three datasets analysed in this chapter were analysed for enrichment on TKIi transcripts using the hypergeometric distribution. Additionally, the existence of two cohorts of datasets transcriptionally profiling TKI responders and non-responder CML patients allowed for

testing the TKI signature as a classifier for TKI response as one dataset was used for training the classifier while the other was used for testing it.

4.2 The TKI signature is able to predict TKI response in CML

4.2.1 *Description of the TKI response microarray dataset*

The RNRMW (responders – non-responders McWeeney) dataset comprises of two different groups of patients: a training group (n=36, peripheral blood) and a validation group (n=23, bone marrow). In both groups there are patients who respond to IM treatment (training: n=24, validation: n=17) and patients who do not (training: n=12, validation: n=6). Patients were defined as responders if they reached complete cytogenetic response (CCyR, 0% Ph⁺ metaphases) after 1 year of IM treatment and as non-responders if they did not achieve even minor cytogenetic response (at least 66% Ph⁺ metaphases) during that time. Affymetrix Human Genome U133 Plus 2.0 chips were used for detecting gene expression in all samples.

PCA was used to summarise the variability of the two datasets present in RNRMW (training and validation) in order to identify batch effects, that would require correction before pursuing with the analysis, and any specific clustering based on a single factor. No clusters associated with TKI response were observed in the projection of the first two PC neither on the training dataset (Figure 4-1B) nor in the validation set (Figure 4-2B). This was confirmed by calculating if there was any significant difference between the values of each eigen vector based on the TKI response classification using Kruskal-Wallis test in both the training (Figure 4-1C) and validation datasets (Figure 4-2C). Small clusters could be observed but the lack of additional clinical information describing the samples prevented further investigation of this observation.

4.2.2 *Replication of the analysis conditions of RNRMW failed to return the same results than presented in the publication*

The original publication (McWeeney et al., 2010) stated that the training set was normalized using RMA and each probe set was compared between the two groups (responders and non-responders) using analysis of variance. The p-values were corrected for false discovery rate (FDR) and filtered for a q-value smaller than 0.1. Additionally, the probes sets were filtered for those with an absolute fold-change greater than 1.5. Using those rules for filtering the gene list after comparing the TKI responders and non-responders with *limma* (Smyth, 2004). No probe sets were found in the training dataset matching this criteria (Figure 4-3A) and only one probe set, which mapped to *HOXA1*, was found to be match this criteria in the validation set (Figure 4-3B).

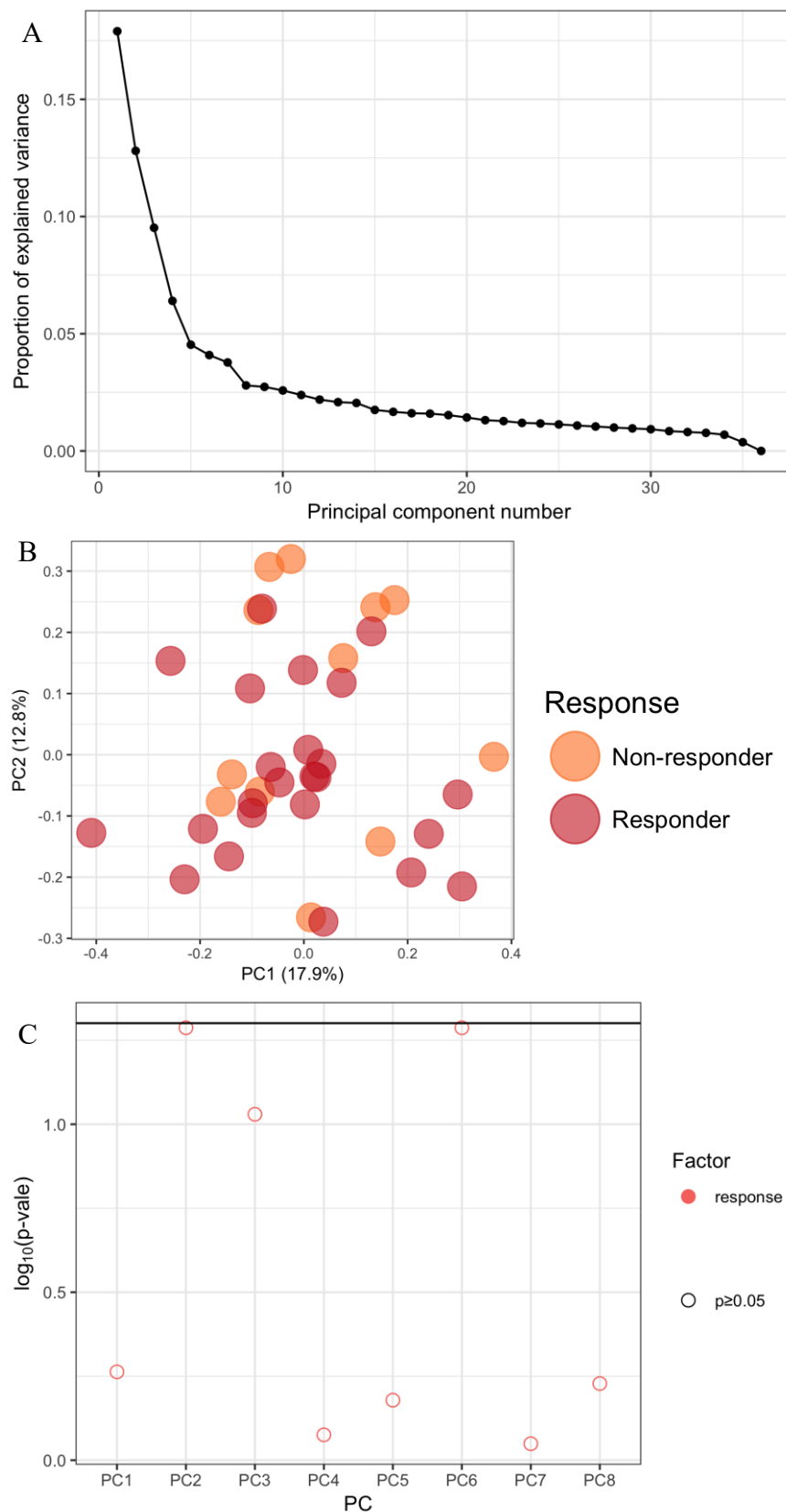


Figure 4-1. TKI response is not associated with any of the first 8 PCs in the training set. The scree plot (A) shows that multiple principal components have an important relative weight in explaining the variability. (B) The projection of PC 1 and 2 fails to discriminate between TKI responders and non-responders. (C) Analysis of the association of the eigen vectors with the response to TKI revealed no significant differences between groups. The black horizontal line marks the threshold $p < 0.05$.

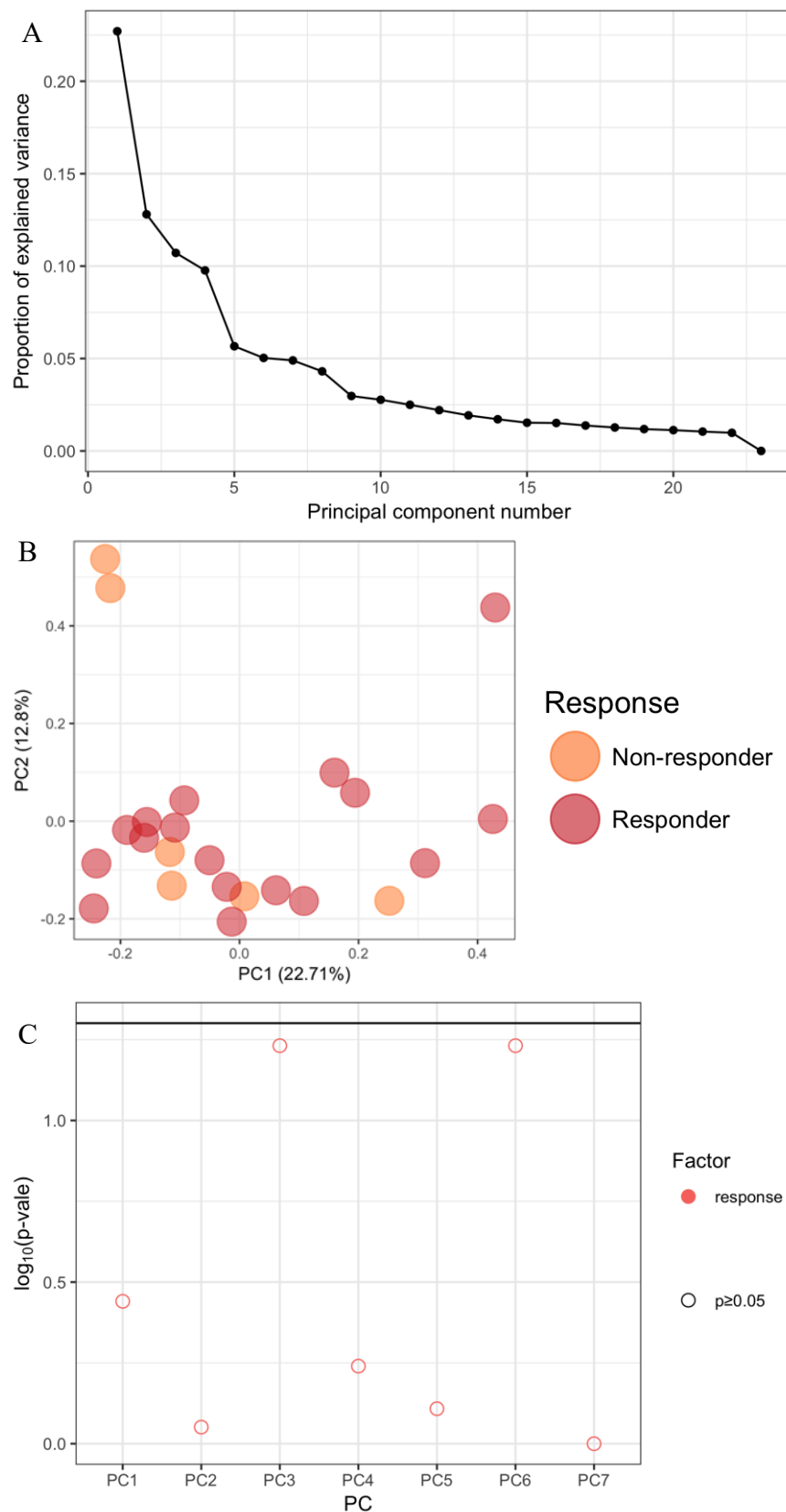


Figure 4-2. TKI response is not associated with any of the first 7 PCs in the validation set. The scree plot (A) shows that multiple principal components have a significant weight in explaining the variability. (B) The projection of PC 1 and 2 fails to discriminate between TKI responders and non-responders. (C) Analysis of the association of the eigen vectors with the response to TKI revealed no significant differences between groups. The black horizontal line marks the threshold $p=0.05$.

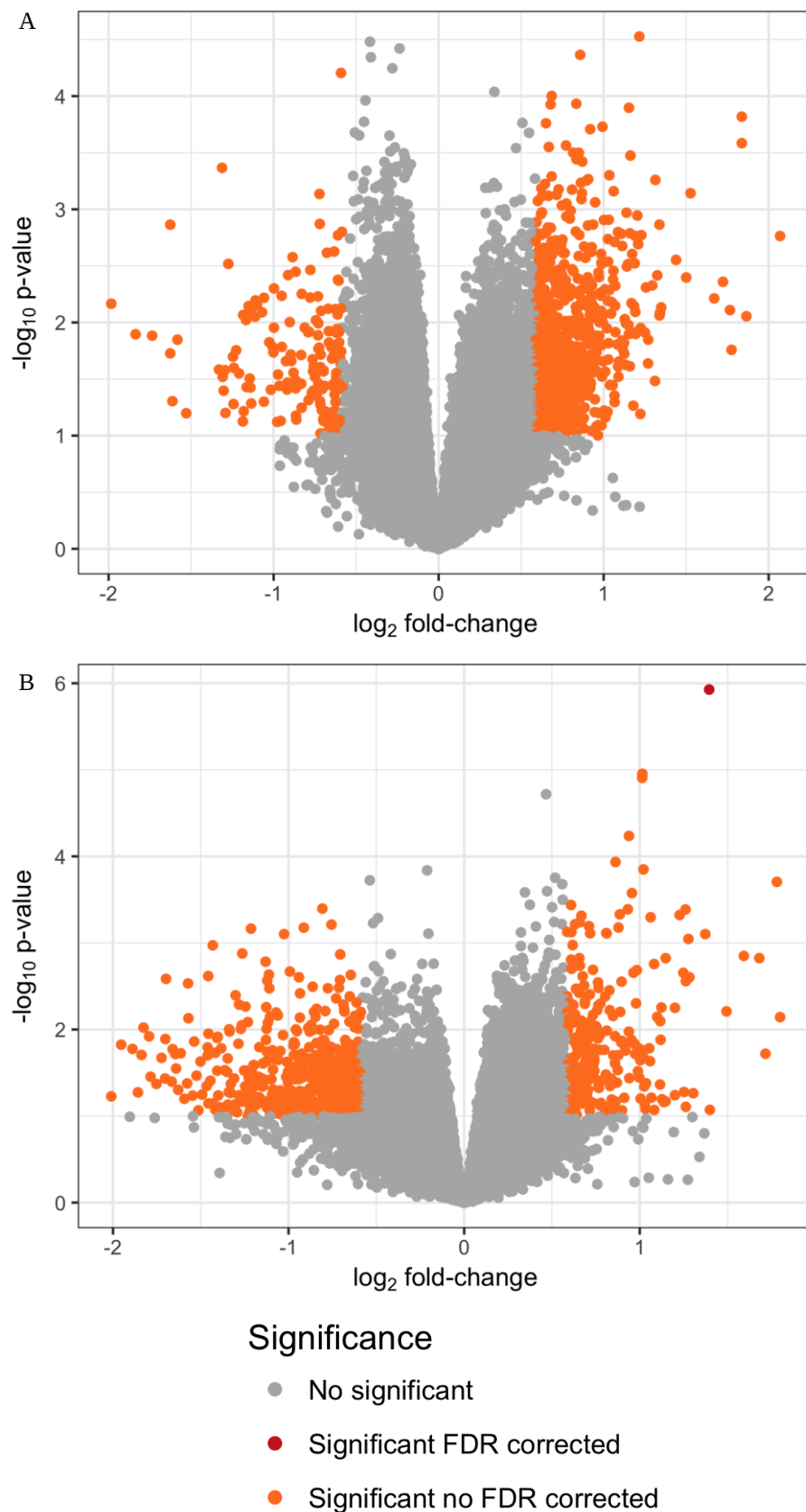


Figure 4-3. Correction for FDR reduces the number of differentially expressed probe sets in the validation set to only *HOXA1*. Probe sets differentially expressed before FDR correction are shown in orange and those which were differentially expressed after correction are shown in red. (A) The training dataset did not present any differentially expressed probe set after FDR correction. (B) A probe set mapping to *HOXA1* was found to be differentially expressed after FDR.

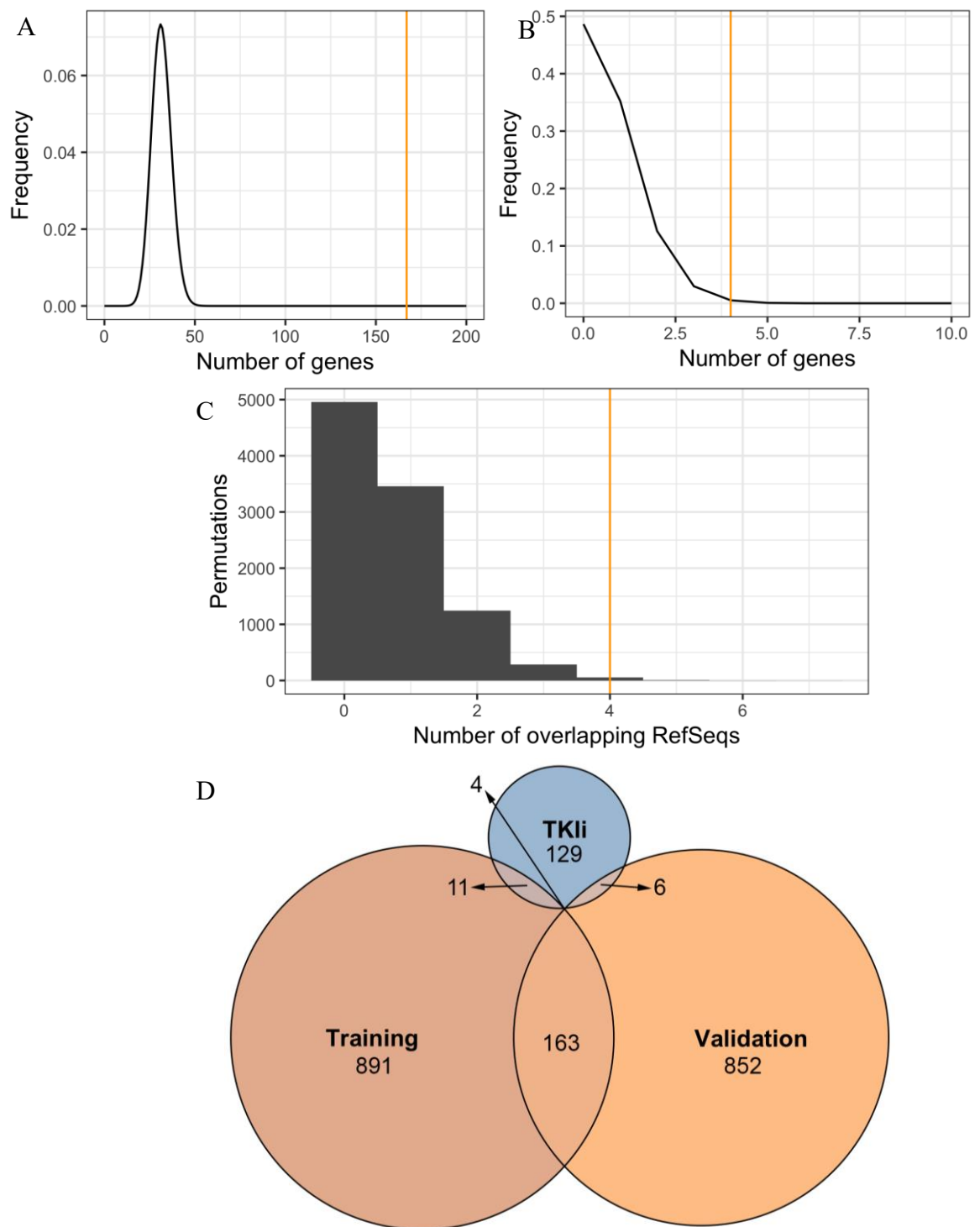


Figure 4-4. The overlap of the differentially expressed RefSeq transcripts is unlikely to happen by chance. Using the hypergeometric distribution it was clear that it is very unlikely for the overlaps between (A) the training and the validation sets and (B) the RNRMW combination and the TKIi₁₆₁ signature to happen by chance. (C) Selecting 10,000 random sets of RefSeqs of the same size than the TKIi signature had a similar result than the hypergeometric distribution. (D) Venn-diagram showing the number of RefSeqs differentially expressed in each set of patients and the overlaps between them and the TKIi signature. The orange vertical bars denote the observed values.

4.2.3 Failing to correct for false discovery rate reproduce similar results to the analysis presented in the original publication

The absence of differentially expressed genes when applying the filters stated in the original publication suggested that it was not possible to replicate the exact methodology from the original publication. This could be due to different normalisation protocols and the different versions of the software used in the two analyses (as the two different analyses have been performed 8 years apart). In order to minimise the differences between the two analyses the filtering for statistical significance was performed with the raw p-value and not by the FDR. Using these settings 938 probe sets were found differentially expressed, a number very close to the one stated by the authors (885) (Figure 4-3A). The similarity between the number of differentially expressed probe sets in both lists increased the confidence in the possibility of replicating the results of original publication with the new criteria.

4.2.4 The overlap between the TKIi signature and the genes differentially expressed between TKI responders and non-responders is bigger than the overlap expected by chance

Having a list of differentially expressed probe sets between TKI responders and non-responders in the training set without correcting for FDR (and using a relaxed p-value threshold of 0.1) would be a more relaxed threshold than the common $q < 0.05$ threshold. The existence of the validation set, however, allowed working with the probe sets differentially expressed in both datasets (training and validation). Using the overlap of the two sets decreases the risk of false positives and increases the confidence in the result. Carrying out the same analysis of the validation set returned 797 differentially expressed probe sets (Figure 4-3B); 87 of these were also differentially expressed in the training set.

In order to investigate if the differentially expressed probe sets were enriched in the TKIi signature, the probe sets were mapped to RefSeq transcripts using *biomaRt* (Durinck et al., 2009). As each probe set can map to a different number of RefSeq transcripts, the numbers of differentially expressed RefSeq transcripts can differ from the number of differentially expressed probe sets. After mapping, 1,069 transcripts were considered differentially expressed in the training set and 1,025 in the validation (Figure 4-4D). Of those, 167 transcripts were differentially expressed in both sets (Figure 4-4D). The probability of finding 167 or more common transcripts in two lists of 1,069 and 1,025 transcripts selected from a pool of 34,769 (number of RefSeq transcripts that map to the probe sets of the microarray) by chance was very small ($p < 1 \times 10^{-70}$, Figure 4-4A). Thus, it was believed that

those RefSeq transcripts are truly differentially expressed between IM responders and non-responders.

The analysis of CMLDV (3.2.1.2), CMLMC (3.2.1.3) and TKIFP (3.2.1.1) datasets described in the previous chapter was repeated to perform it at the transcript level (instead of summarising the probe sets to the gene level). This new analysis reported 161 TKI transcripts, of which 150 had a probe set mapping to them in the Affymetrix Human Genome U133 Plus 2.0 chip (used in RNRMW). The TKI₁₆₁ list was compared to the list of differentially expressed transcripts obtained from the overlap of the two sets of RNRMW, finding 4 common transcripts that coded for EGFL6, VWF, CACNA1D and RBPM5. The probability of finding an overlap of 4 or more transcripts between the two lists by chance was found to be unlikely by both hypergeometric distribution ($p=0.0059$; Figure 4-4B) and Monte-Carlo by selecting 10,000 random sets of 150 RefSeq IDs ($p=0.0062$; Figure 4-4C).

4.2.5 The TKI signature can discriminate between TKI responders and non-responders

The significant overlap between the TKI-response transcripts and the TKI₁₆₁ signature suggested that the TKI₁₆₁ signature might actually have potential for predicting the TKI response a patient will have. The projection of the first two PCs of the TKI₁₆₁ signature revealed a better separation of TKI responders and non-responders than when the PCA was done with all the transcripts in both the training (Figure 4-5B) and the validation datasets (Figure 4-6B). This was also confirmed by the association of the TKI response with the PC1 in the training dataset (Figure 4-5C) and with PC2 in the validation dataset (Figure 4-6C). However, the separation in different clusters was not as clear as with the 938 probe sets that were differentially expressed in the training dataset (Figure 4-7B), which showed association of the TKI response with the PC1 (Figure 4-7C), but not in the validation dataset (Figure 4-8B), which showed association of the TKI response with the PC4 (Figure 4-8C). The 75 probe sets of the classifier built in the original publication improved the separation in different clusters based on the response to TKI in both the training dataset (Figure 4-9B) and the validation dataset (Figure 4-10B) associating in both datasets with the PC1 (Figure 4-9C and Figure 4-10C). In summary, The TKI₁₆₁ signature improved the clustering of the samples in TKI responders and not responders but was not as successful as the genes present in the classifier built in the original publication.

This modest but positive result at discriminating the response to TKI using the TKI₁₆₁ signature suggested that the TKI transcripts could be used to build a classifier to predict

TKI response. A Support Vector Machine (SVM) classifier was chosen based on its consistent high performance across a wide range of domains. The scikit-learn SVM model was used with default options (RBF kernel) (Pedregosa et al., 2011). The existence of two distinct datasets in RNRMW allowed training the classifier in the “training” dataset and testing it in the “validation” dataset. The classifier built with the TKI₁₆₁ signature had an AUC of 0.82 in the RNRMW’s validation dataset (Figure 4-11). This result was better than the result obtained using *all* available probe sets in the universe (AUC 0.73) and only 2.9% of classifiers built using random subsets of transcripts of the same size than the TKI₁₆₁ signature (150) achieved an equal or better result, implying $p < 0.05$. However, the 149 RefSeq transcripts mapped to the 75 probe sets of McWeeney’s classifier got an AUC of 0.87, which was better than the TKI₁₆₁ signature and only 1.0% of the random subsets perform equally or better than it (Figure 4-11). This improvement in AUC and statistical significance of the McWeeney’s classifier was expected as the classifier was optimised (e.g. selection of optimal number of probe sets) and built using this dataset. The achievement of the TKI₁₆₁ signature as classifier is that a list of transcripts generated from a completely different research question (transcripts differentially expressed in CML but not affected by TKI treatment) has a significantly better AUC than random sets of transcripts.

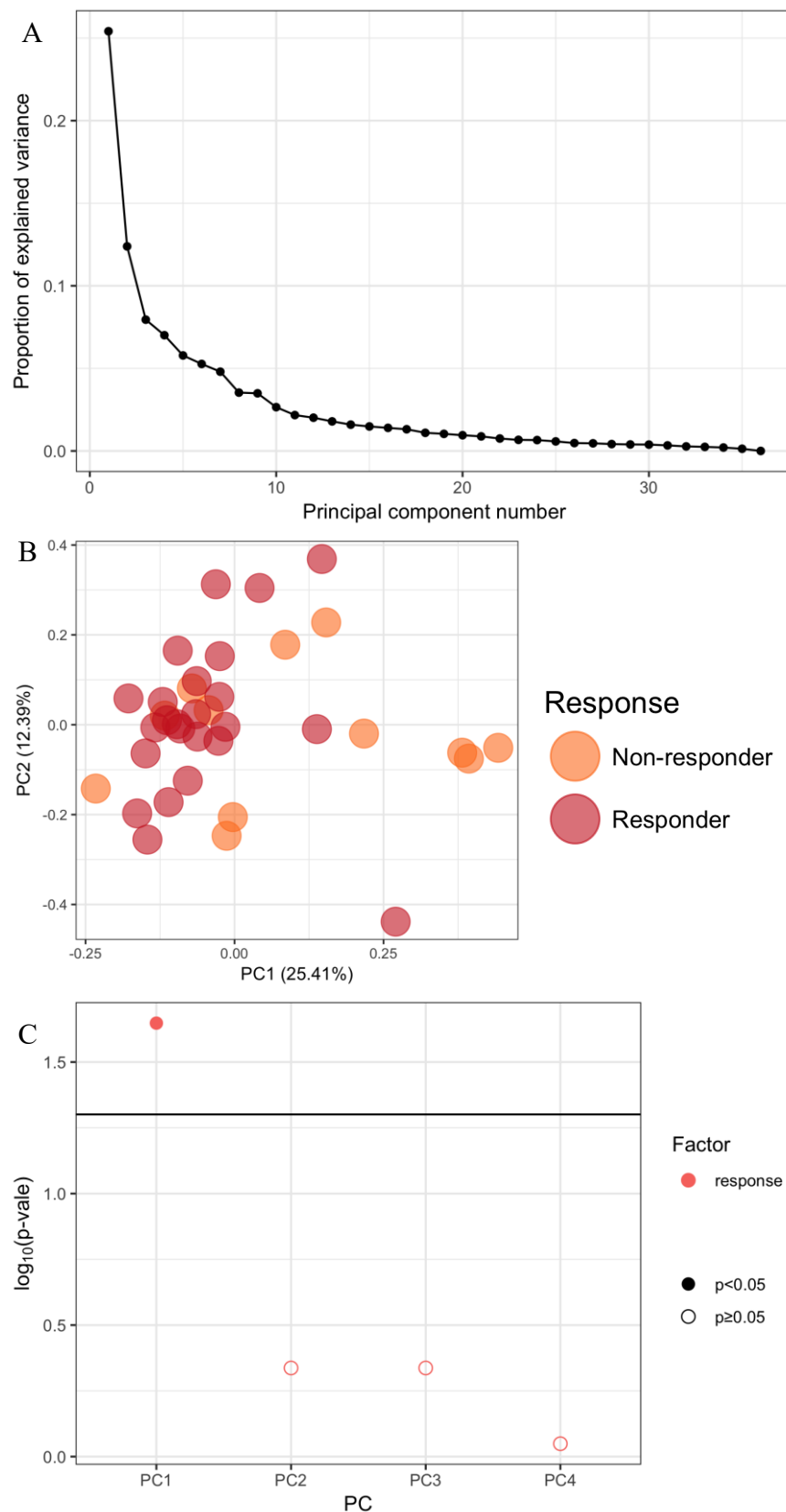


Figure 4-5. Selecting only the TKI₁₆₁ RefSeqs increases the discrimination between responders and non-responders in the training set. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. It can be observed that TKI responders and non-responders separate more than with all the probe sets. (C) Analysis of the association of the eigenvectors with the response to TKI revealed an association between the PC1 and the response to TKI. The black horizontal line marks the threshold $p < 0.05$.

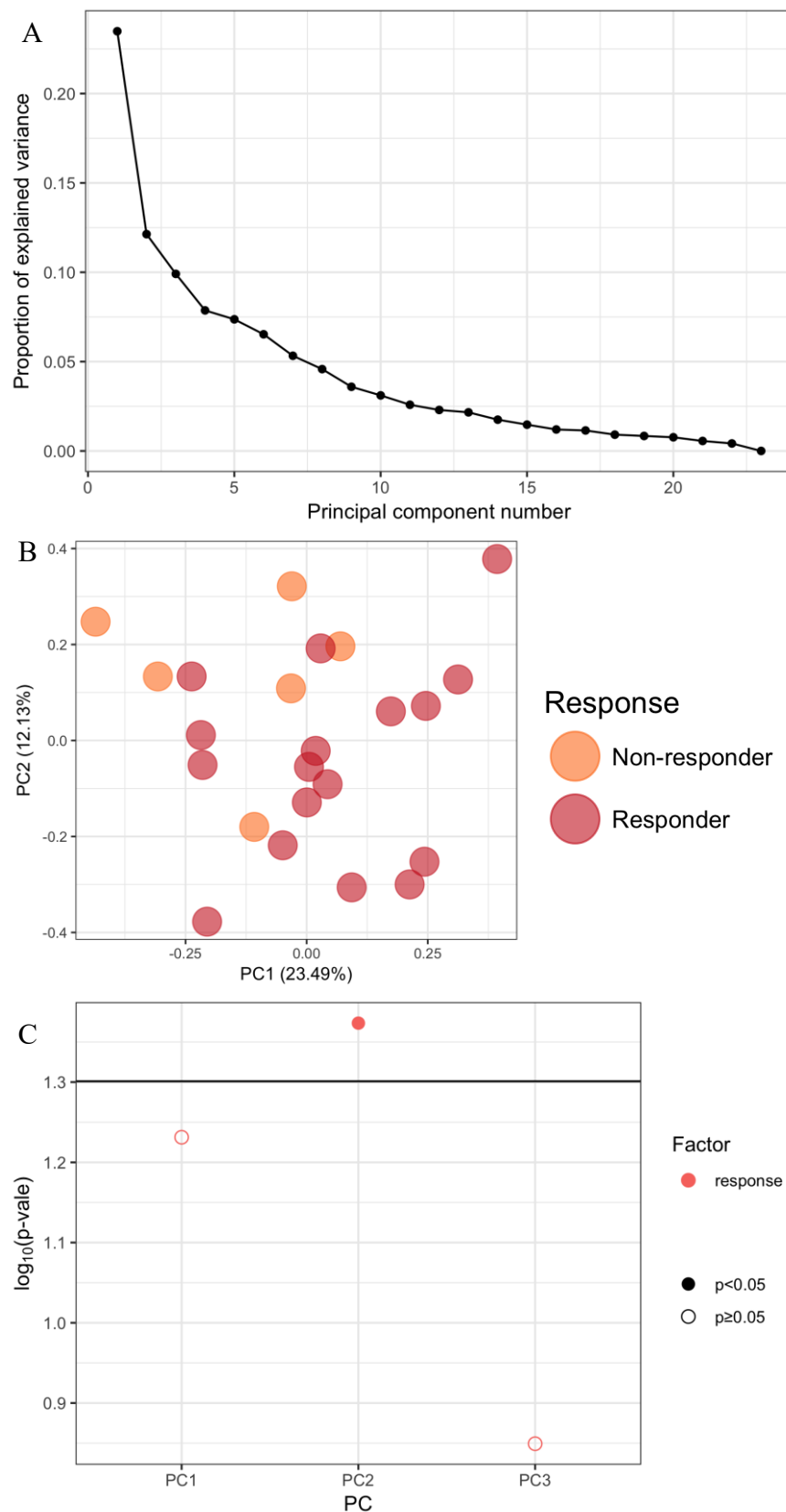


Figure 4-6. Selecting only the TKI₁₆₁ RefSeqs increases the discrimination between responders and non-responders in the validation set. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. It can be observed that TKI responders and non-responders separate more than with all the probe sets. (C) Analysis of the association of the eigen vectors with the response to TKI revealed an association between the PC2 and the response to TKI. The black horizontal line marks the threshold $p < 0.05$.

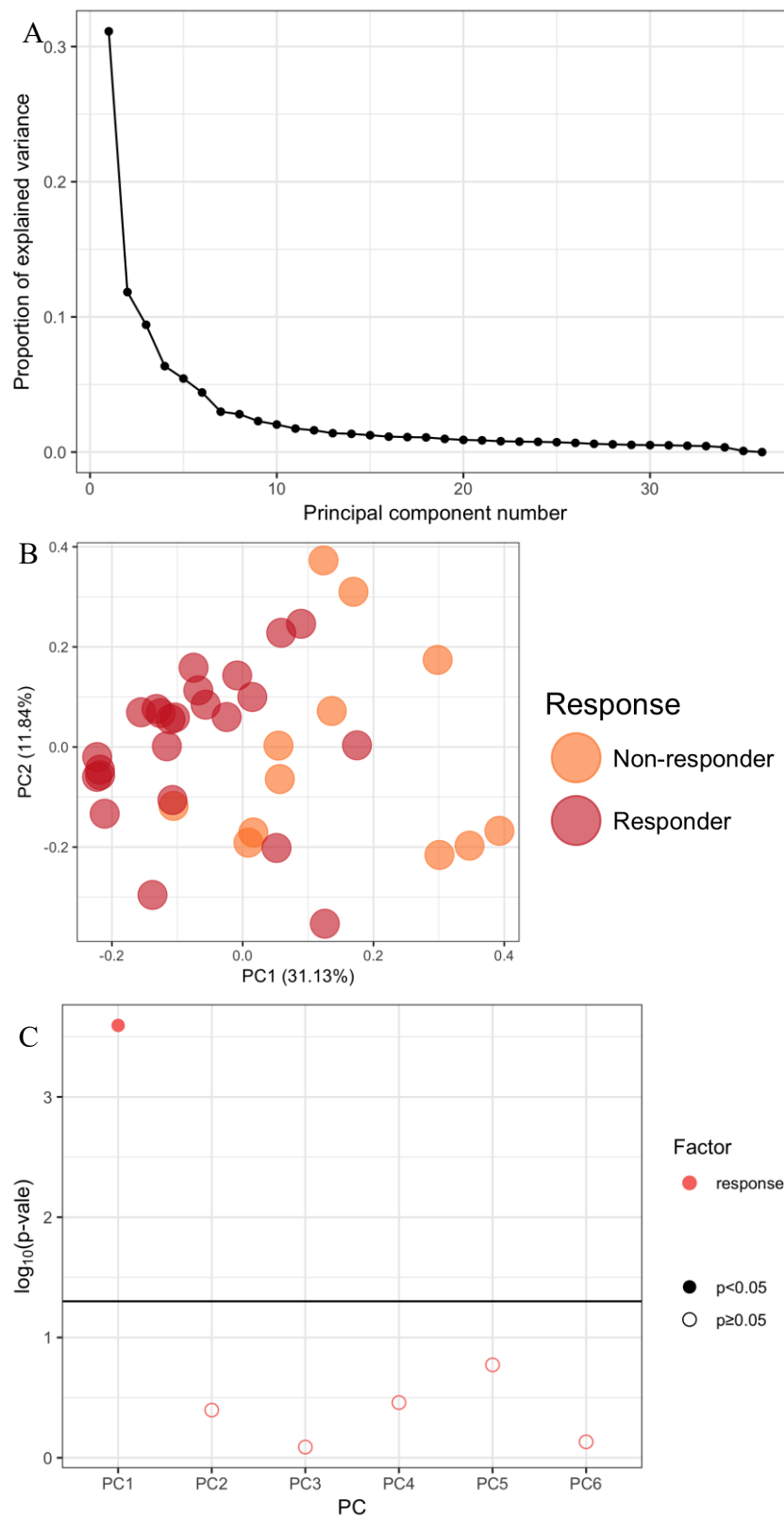


Figure 4-7. Selecting only for the probe sets differentially expressed in the training set increases the discrimination between responders and non-responders in the training set. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. It can be observed that TKI responders and non-responders separate more than with all the probe sets. (C) Analysis of the association of the eigen vectors with the response to TKI revealed an association between the PC1 and the response to TKI. The black horizontal line marks the threshold $p < 0.05$.

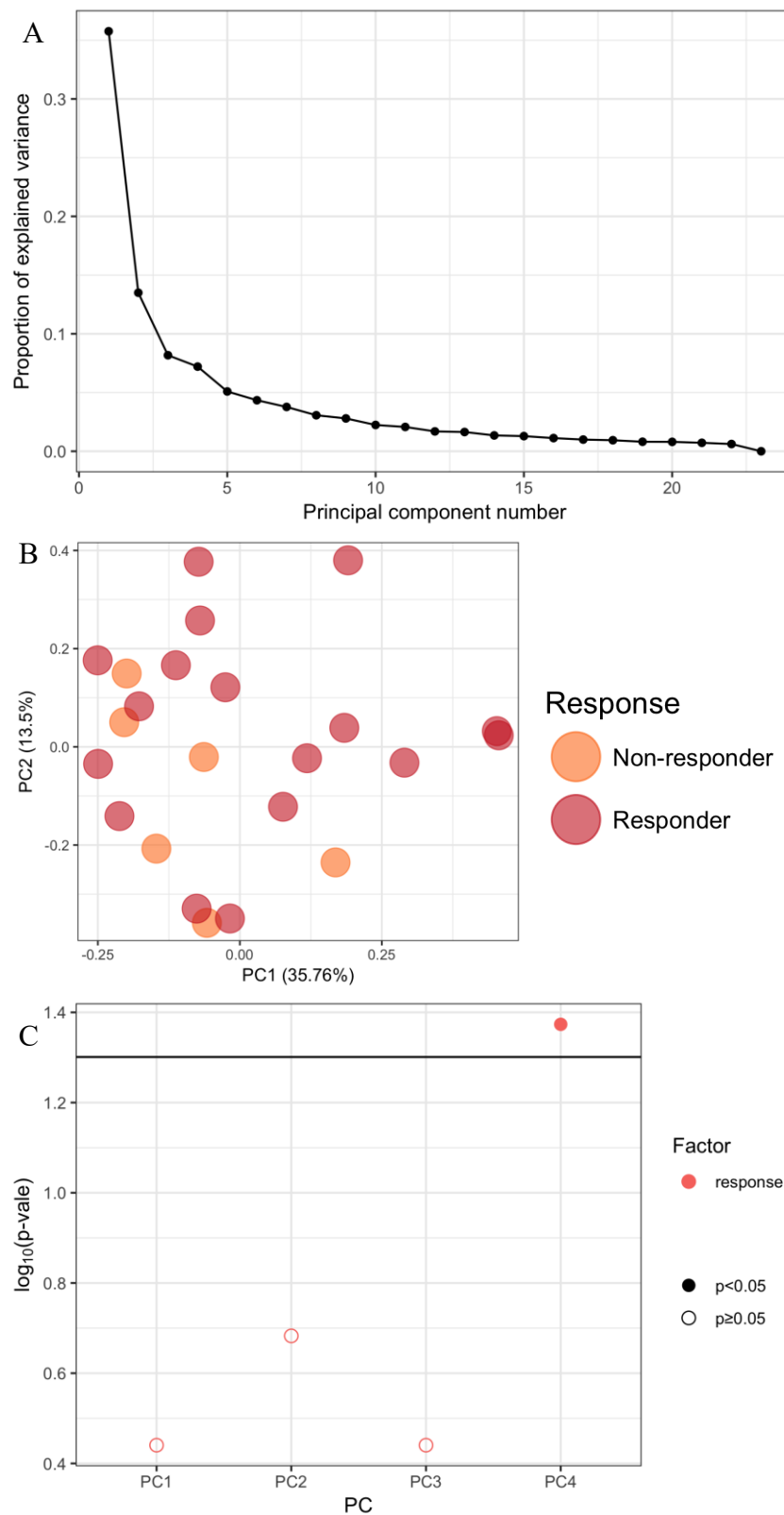


Figure 4-8. Selecting only for the probe sets differentially expressed in the training set increases the discrimination between responders and non-responders in the validation set. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. It can be observed that TKI responders and non-responders modestly separate more than with all the probe sets. (C) Analysis of the association of the eigen vectors with the response to TKI revealed an association between the PC4 and the response to TKI. The black horizontal line marks the threshold $p < 0.05$.

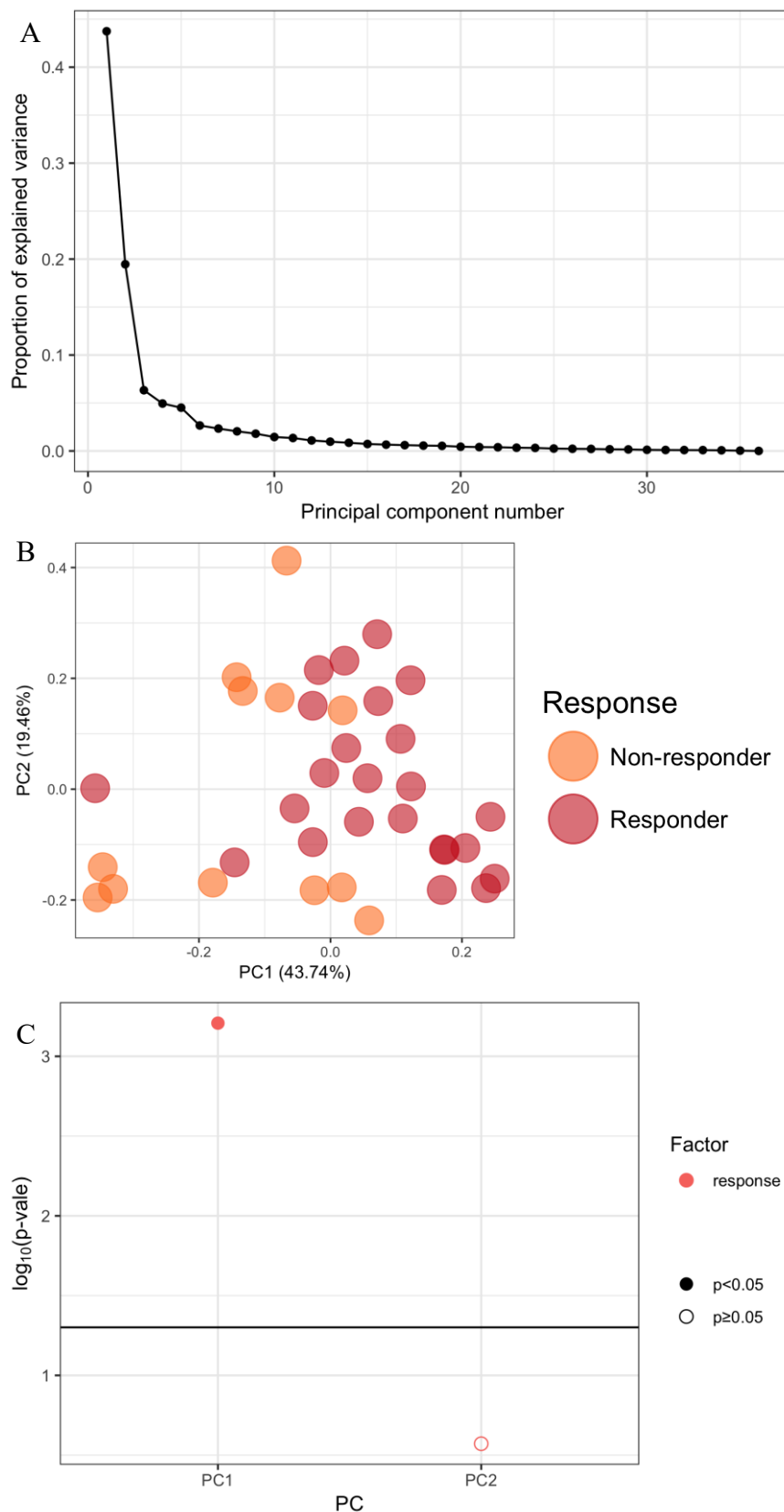


Figure 4-9. Selecting only for the probe sets present in McWeeney's classifier has the largest effect in discriminating between responders and non-responders in the training set. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. It can be observed that TKI responders and non-responders separate more than with all the probe sets. (C) Analysis of the association of the eigen vectors with the response to TKI revealed an association between the PC1 and the response to TKI. The black horizontal line marks the threshold $p < 0.05$.

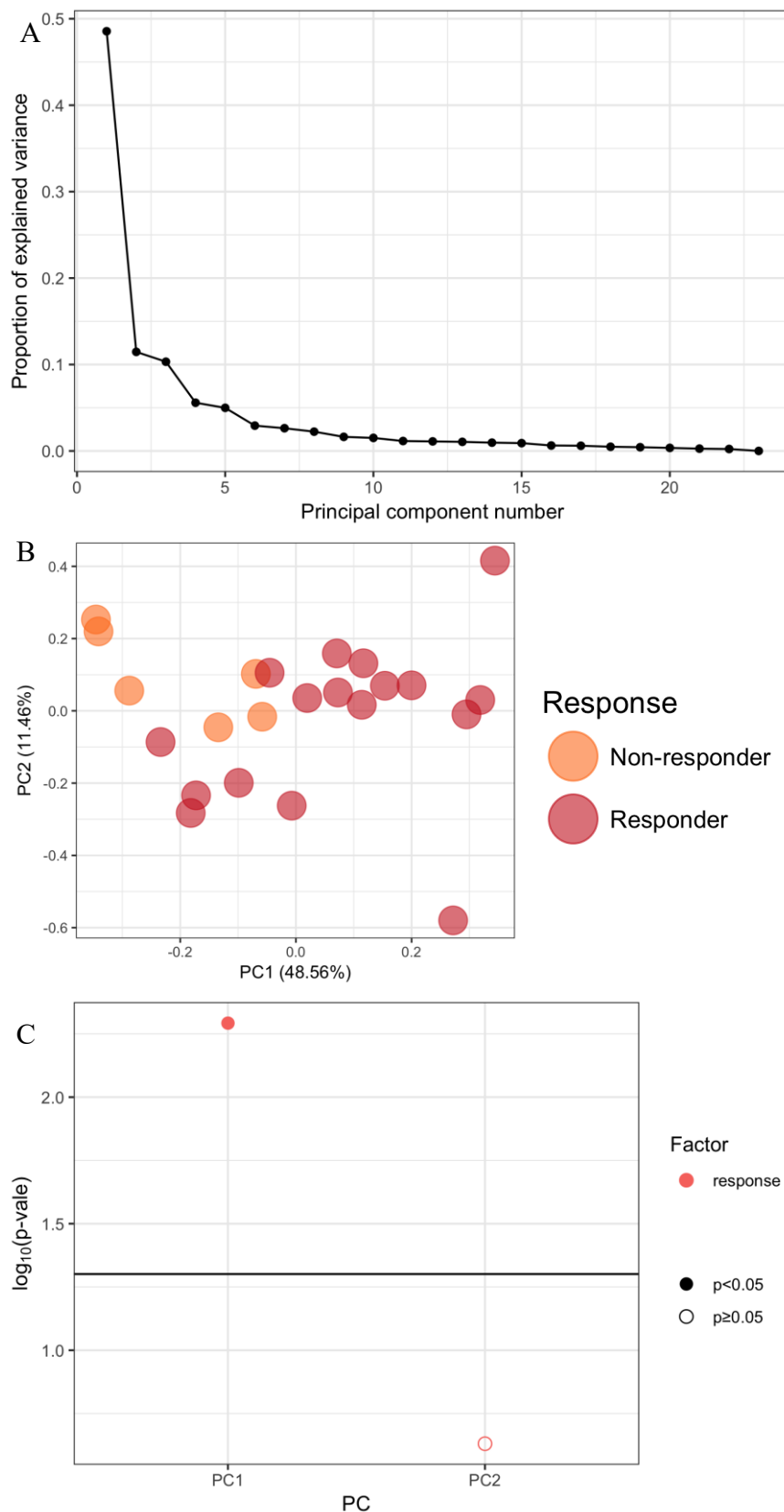


Figure 4-10. Selecting only for the probe sets present in McWeeney's classifier has the largest effect in discriminating between responders and non-responders in the validation set. (A) Percentage of variance explained by each PC. **(B)** Projection of the first two PCs. It can be observed that TKI responders and non-responders separate more than with all the probe sets. **(C)** Analysis of the association of the eigen vectors with the response to TKI revealed an association between the PC1 and the response to TKI. The black horizontal line marks the threshold $p < 0.05$.

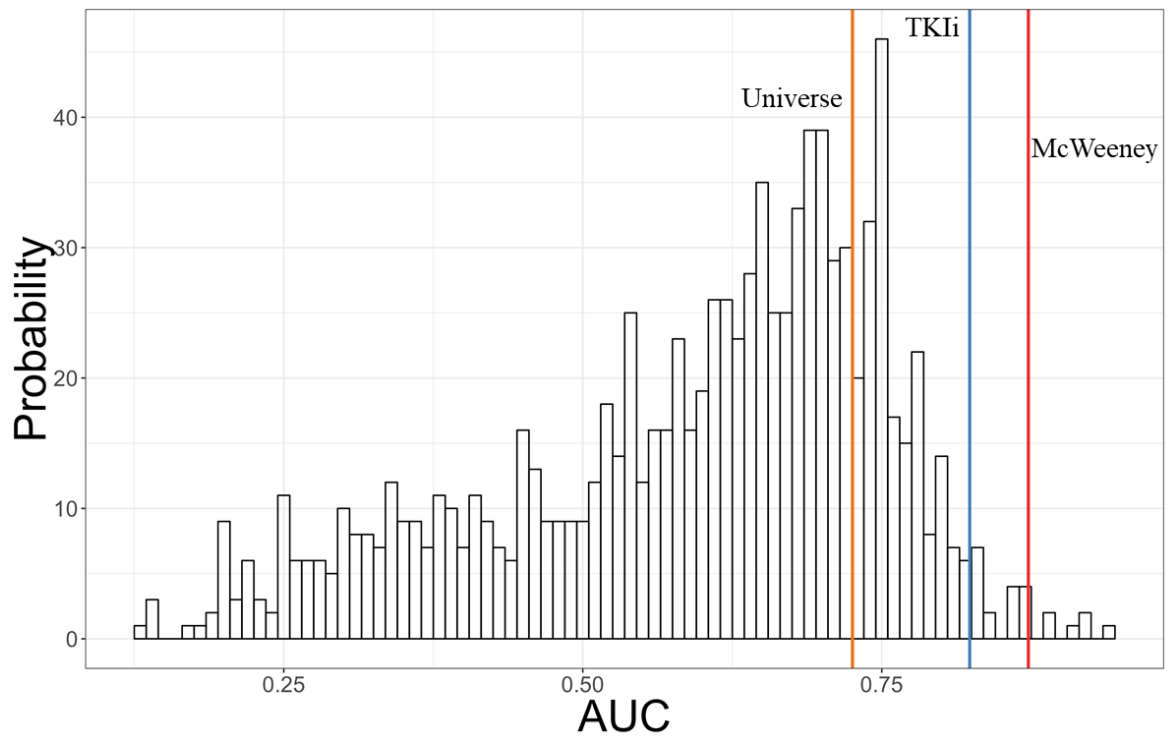


Figure 4-11. The TKLi₁₆₁ signature and the McWeeney classifier have higher AUCs than random. The histogram shows a distribution of random sets of 150 probe sets (same size than the TKLi₁₆₁ signature in the microarray platform used in RNRMW datasets). The orange bar represents the AUC obtained by using all the probe sets in the universe (0.73, $p=0.21$), the blue bar represents the AUC of the TKLi₁₆₁ signature (0.82, $p=0.029$) and the red bar represents the AUC of McWeeney classifier (0.87, $p=0.01$).

4.3 The TKli signature is not able to predict disease aggressiveness

4.3.1 *Description of the disease aggressiveness microarray dataset*

The aggressive versus indolent dataset generated by Agnes Yong (AIAY) (Yong et al., 2006) was generated using CML patient samples collected between 1979 and 2001 from leukapheresis (peripheral blood) and enriched for CD34⁺ cells (>90%). The different samples were classified as indolent disease when the patients survived at least 7 years before developing a blast transformation (n=9), and as aggressive disease when the patients developed blast transformation within 3 years after diagnosis (n=10). The treatment used on these patients was interferon- α and/or hydroxyurea. Affymetrix Human Genome U133A 2.0 chips were used for detecting gene expression in all the samples.

Projection of the first 2 PCs of the PCA did not reveal clustering based on the aggressiveness of the disease but did reveal clustering based on gender on the PC2 (Figure 4-12B). Study of the association of the different PCs with gender, age and aggressiveness confirmed the association of PC2 with the gender. Additionally, the PC3 associated with the aggressiveness of the disease and PC4 with the age (Figure 4-12C). This suggested that although PC1 was not associated with any of the known factors, gender, aggressiveness and age have an effect in gene expression, as it is shown by their association with the PCs 2, 3 and 4.

4.3.2 *Differential expression analysis revealed no significant differences between aggressive and indolent patients*

Although the original publication (Yong et al., 2006) analysed the gene expression data using three different bioinformatics tools, no FDR was applied. A better comparison of the differentially expressed transcripts of AIAY and RNRMW could be achieved by applying the same normalisation and differential expression protocol. However, AIAY is only one dataset and therefore, there is no option for validating the differentially expressed transcripts in a sister dataset. Because of that, it was decided to apply FDR<0.1 and fold-change>1.5 threshold during *limma* (Smyth, 2004) differential expression analysis for considering a transcript differentially expressed. However, no transcripts met these criteria (Figure 4-13).

4.3.3 *The TKli signature is not a marker for disease aggressiveness*

Although a list of differentially expressed genes was not generated, it was decided to investigate the potential of the TKli₁₆₁ signature for predicting aggressiveness in CML. 103 TKli RefSeq transcripts were present in the AIAY dataset and PCA was carried out using only those data. Kruskal-Wallis test showed that none of the principal components was

significantly associated with any of the known factors; although aggressiveness and gender were almost significantly associated with principal components 2 and 3 respectively (Figure 4-14B-C).

Performing PCA using only the 57 probe sets that match to the probe sets included in the McWeeney classifier (McWeeney et al., 2010) increased the variance explained by PC1 to 45% (Figure 4-15A), which associates significantly with both aggressiveness and age (Figure 4-15B-C). This increase in the variance significantly associated with either aggressiveness or the age group (from 16.2% to 45%) suggests that using the McWeeney classifier reduces the noise for discriminating between patients with different disease aggressiveness and different age groups. However, the association of both factors with PC1 in this analysis was equally significant, which may suggest that both factors are co-founders in the gene expression changes observed in the genes present in the McWeeney classifier.

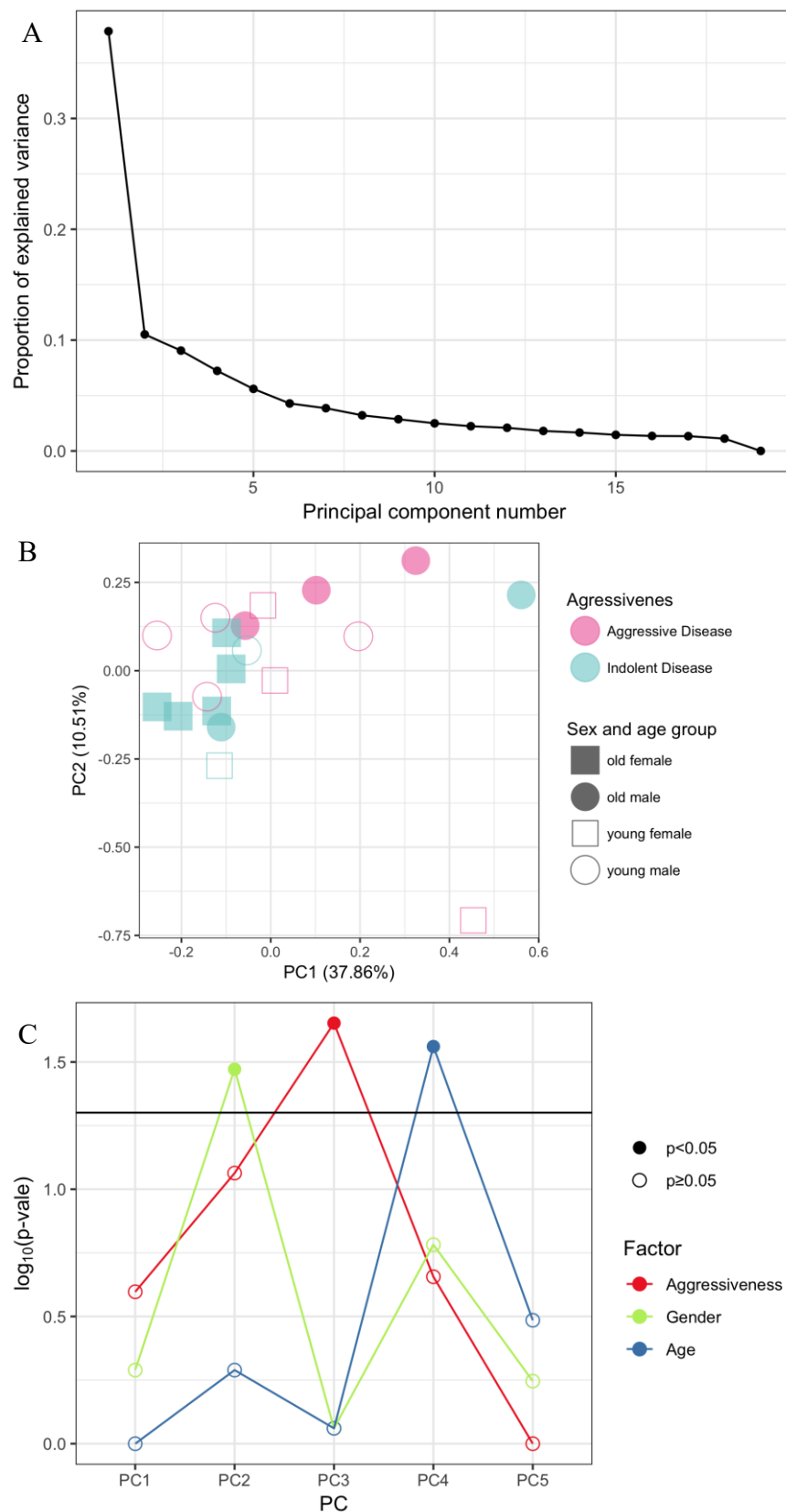


Figure 4-12. AIAY's PCA reveals association of gender, aggressiveness and age with PCs 2, 3 and 4. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. PC2 is associated with a separation between male and female patients. (C) Analysis of the association of the eigen vectors with the known factors (gender, aggressiveness and age) revealed an association between them and the PCs 2, 3 and 4. The black horizontal line marks the threshold $p < 0.05$.

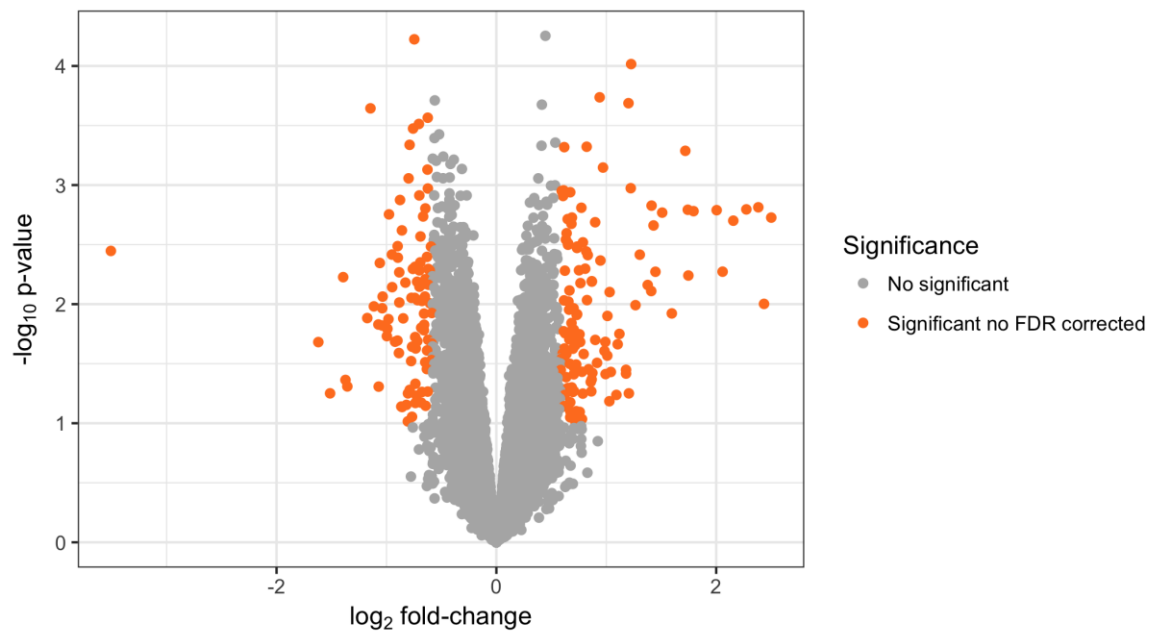


Figure 4-13. Differential gene expression analysis revealed no significantly DE genes between aggressive and indolent disease in CML. The probe sets that had a non-adjusted $p < 0.1$ and a fold change > 1.5 are highlighted in orange.

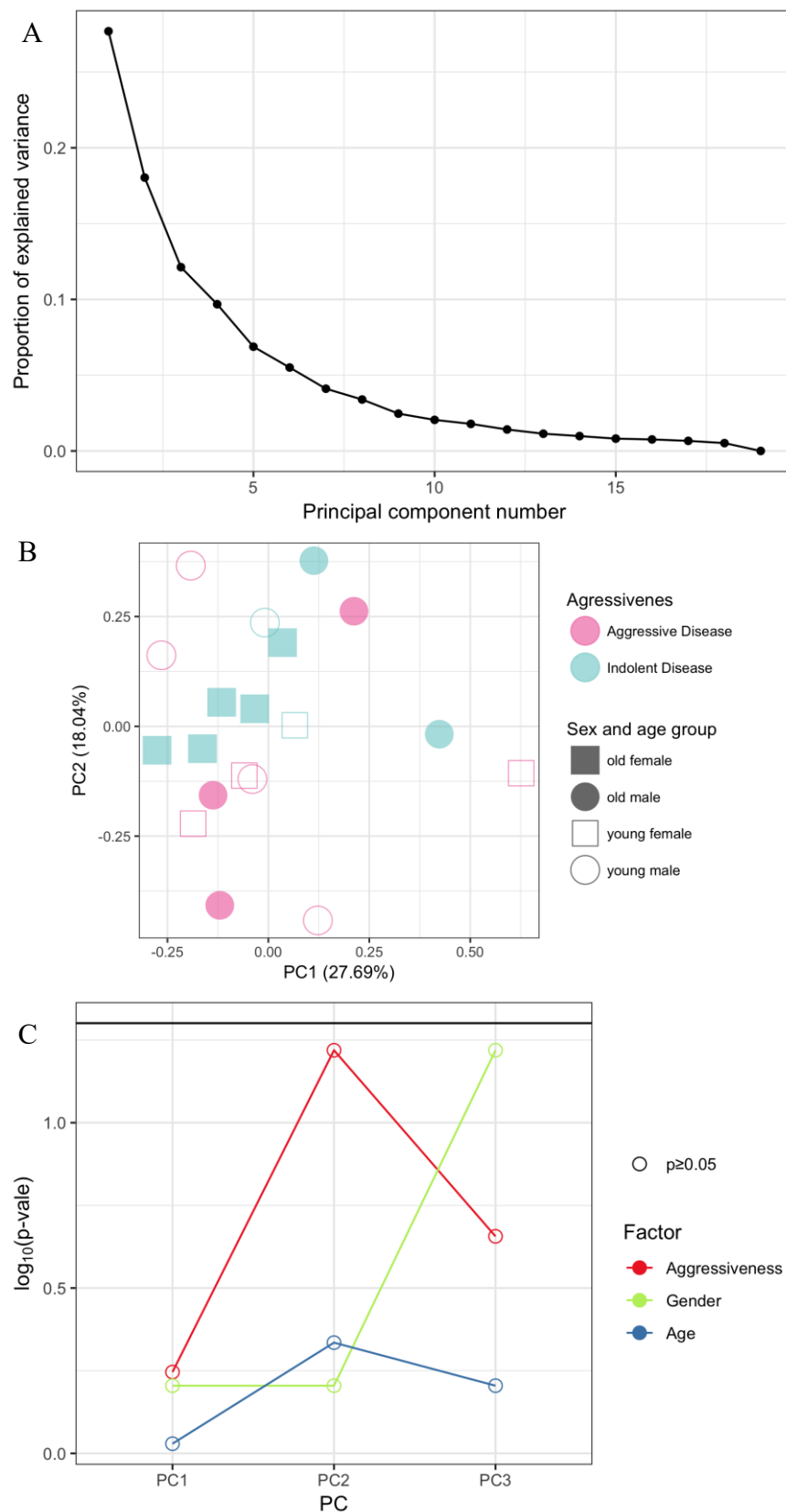


Figure 4-14. Selecting only for the TKLi₁₆₁ transcripts reduced the association of the PCs with any of the known factors (disease aggressiveness, age and gender). (A) Percentage of variance explained by each PC. **(B)** Projection of the first two PCs. No clear clustering was observed. **(C)** No associations were found between the eigen vectors and the known factors (gender, aggressiveness and age). The black horizontal line marks the threshold $p < 0.05$.

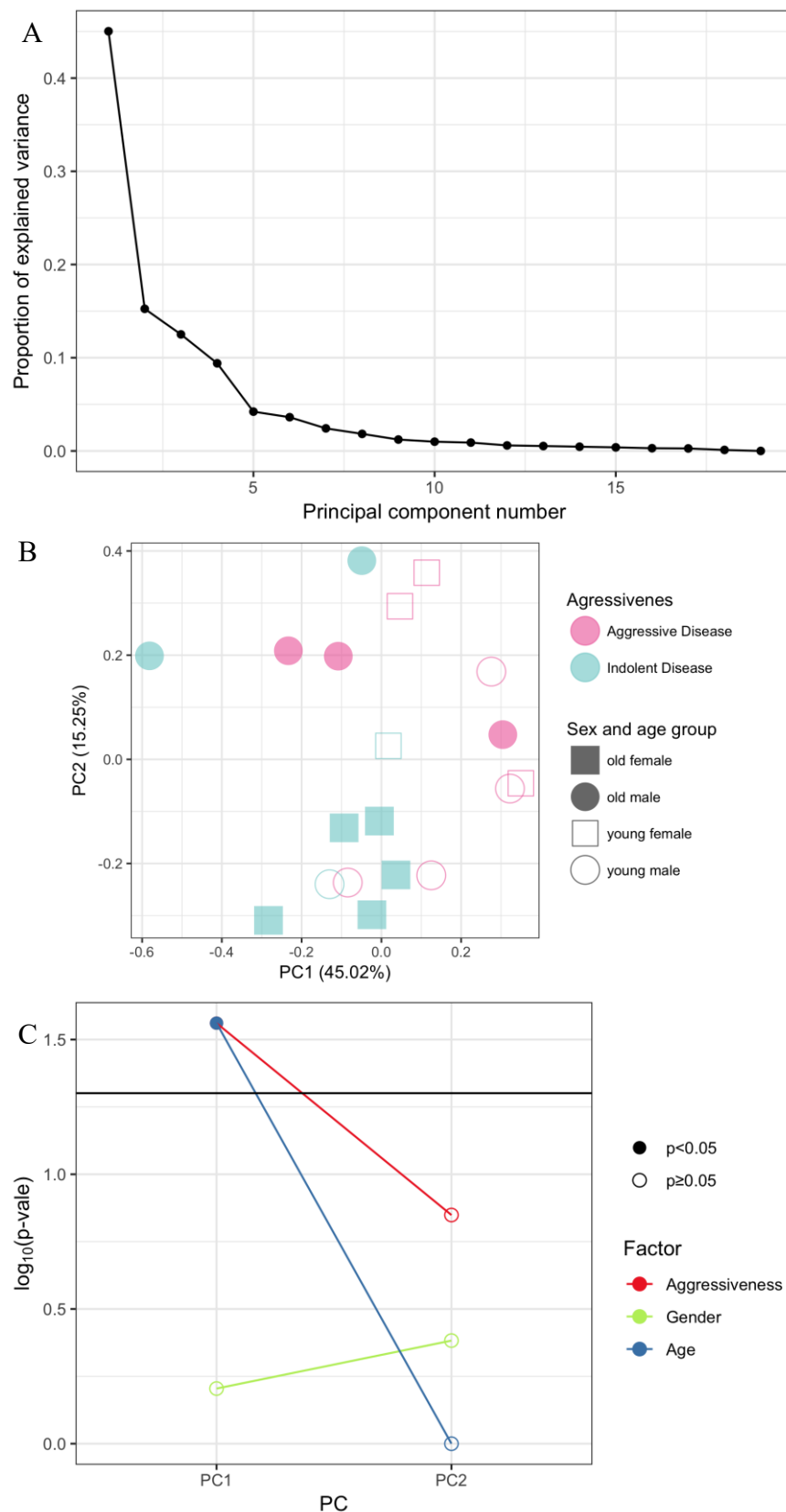


Figure 4-15. Selecting only for the McWeeney classifier transcripts associated the PC1 with disease aggressiveness and age. (A) Percentage of variance explained by each PC. (B) Projection of the first two PCs. PC1 is associated with a separation between old and young patients as well as those suffering from an indolent or aggressive disease. (C) PC1 was associated with the known factors aggressiveness of the disease and age. The black horizontal line marks the threshold $p < 0.05$.

4.4 Disease progression is associated with the expression of the TKI signature

4.4.1 *Disease phase and cell population are strongly associated with the variance in the dataset*

In order to assess the role of the TKI₆₀ signature in disease progression, the BC and AP samples from CML Mhairi Copland (CMLMC) dataset were analysed. PCA revealed a stronger association of the cell population than of the disease phase in the first two principal components (Figure 4-16B-C). However, the phase of the disease was also significantly associated with the first six principal components (Figure 4-16C). The significant association across several principal components suggests that disease progression has a strong effect on gene expression.

4.4.2 *Differential gene expression reveals an enrichment on TKI genes in BC and AP*

Differential gene expression analysis was performed using *limma* and a FDR<0.1 was used as threshold for significance. 2,810 genes were found to be differentially expressed in AP HSCs when compared with HSCs from normal controls but only one gene, *KCTD14*, was found to be differentially expressed (upregulated) in AP when compared with CP HSCs. Pathway overrepresentation was studied in the 2,810 differentially expressed genes in AP using PANTHERdb (Mi et al., 2017). This analysis revealed that components of cell cycle, DNA biogenesis and p53 pathways were overrepresented while G-protein signalling pathway components were underrepresented (Fisher's exact test FDR<0.05; Table 4-1).

The same analysis was performed in BC HSCs. When compared with normal HSCs, it was found that 3,707 genes were differentially expressed in BC. However, no pathways were found to be over or underrepresented as reported by PANTHERdb. 6,280 genes were differentially expressed in BC when compared with CML CP HSCs. Using PANTHERdb DNA replication, apoptosis and p53 pathways were found to be overrepresented in the genes differentially expressed in BC when compared with CP (Table 4-2).

The results obtained from the analysis of the AP samples' gene expression are in line with what was expected, as AP cells presented a number of differentially expressed genes when compared with normal controls but were very similar to CP. This is consistent with the positive response to TKI treatment in most patients that evolve into AP (le Coutre et al., 2012), which at the same time may suggest a limited amount of mutations or expression changes compared with CP. The overrepresented pathways are consistent with the increase in proliferation and survival observed in CML cells compared with normal cells. The results from the gene expression analysis of the BC samples revealed more differences

between BC and CP than between BC and normal. The overrepresented pathways (p53, apoptosis, and DNA replication) are consistent with the more aggressive phenotype of the disease (more proliferation and less cell differentiation). However, only two CML BC samples were included in the dataset and therefore the results could change after addition of additional samples.

The overlap between the TKI₆₀ genes and these lists of differentially expressed genes was calculated and assessed. The 2,810 genes differentially expressed between AP and normal contained 48 of the 60 TKI₆₀ genes, the genes differentially expressed between BC and normal contained 36 TKI₆₀ genes and, the comparison of BC and CP contained 16 TKI₆₀ genes. All of these overlaps are highly unlikely to happen only by chance when assessed by hypergeometric distribution ($p < 0.001$ in all cases; Table 4-3).

4.4.3 Selecting only the TKI genes increases the discrimination between CML phases

Selecting only the 60 TKI₆₀ genes on the whole dataset (including all phases and cell populations) increased the association of the disease phase and disease while decreasing its association with the cell population when performing PCA (Figure 4-17C) compared with the PCA performed using all the probes on the chip (Figure 4-16C). This suggested that the relative expression of the TKI₆₀ genes is similar over all the phases of CML when compared with the normal controls as the first principal component, which accounts for 45.88% of the variance (Figure 4-17A), separates mainly CML and normal samples. This was confirmed by observing the relative gene expression, which is consistently up or down regulated compared with the normal controls in the HSCs of the three CML phases (Figure 4-18). Only *CD33*, *GMPR*, *RAB38* and *UBASH3B* had different directions of change in BC compared with CP and AP. CP and AP relative expression of the TKI₆₀ genes was always in the same direction. These four genes could explain why the TKI₆₀ signature also seems to discriminate BC from other phases of the disease on principal component 2, which accounts for 10.93% of the variance (Figure 4-17B).

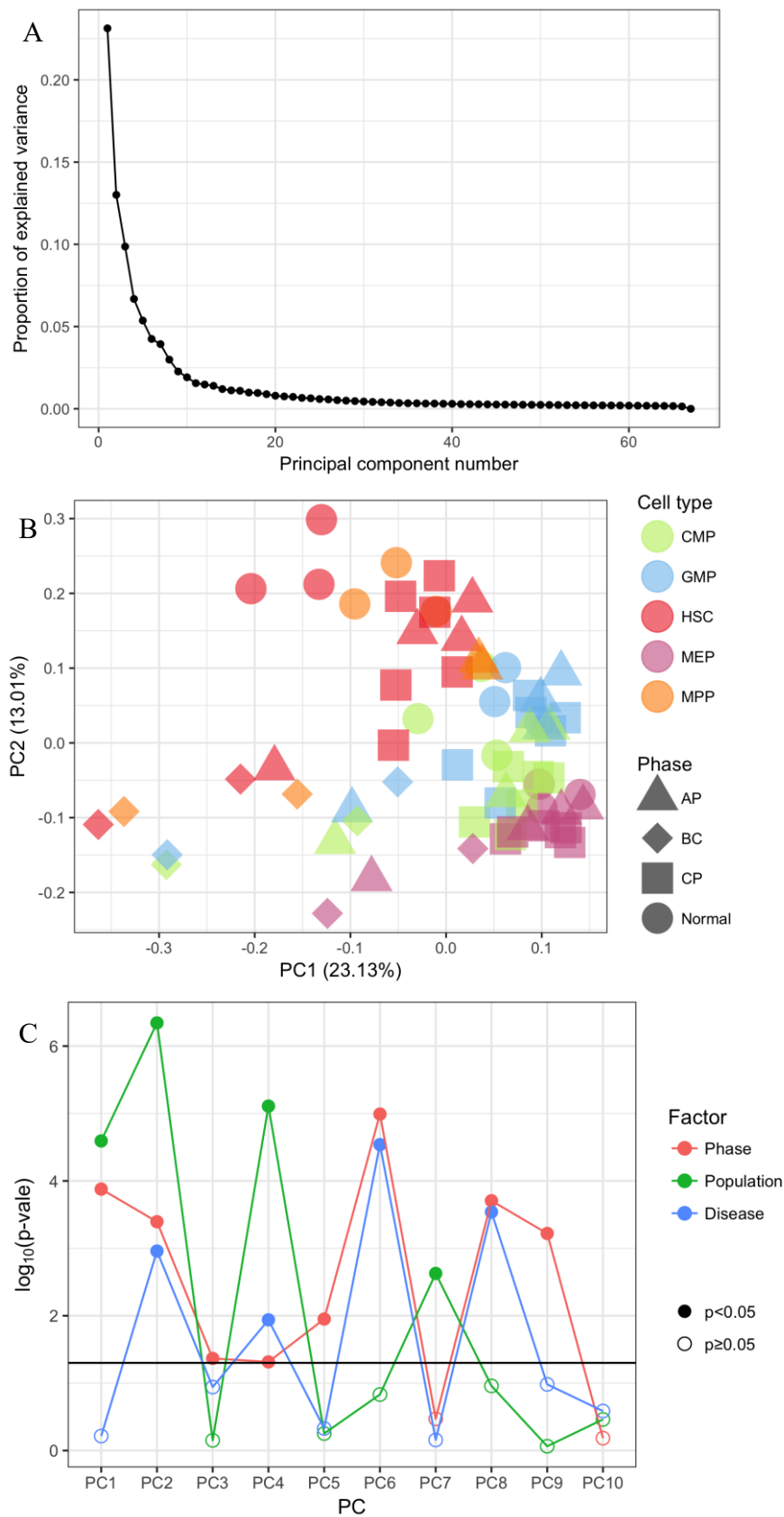


Figure 4-16. Principal components are strongly associated with all the three known factors in CMLMC: cell population, phase of the disease and presence of BCR-ABL1. (A) Percentage of variance explained by each PC. (B) The projection of the first 2 principal components shows clustering by cell type and phase of the disease. (C) The first 9 principal components are strongly associated with at least one know factor. The black horizontal line marks the threshold $p < 0.05$.

Table 4-1. Over and under-represented signalling pathways in the DE genes between AP and normal controls. The genes were mapped to PANTHERdb pathways and enrichment was calculated using the Fisher exact test and BH correction for multiple testing.

PANTHER Pathways	Fold-change	FDR
De novo pyrimidine biosynthesis	6.01	9.97E-03
DNA replication	5.23	3.13E-04
Cell cycle	4.48	8.94E-03
De novo purine biosynthesis	4.41	1.93E-03
Ubiquitin proteasome pathway	3.82	5.20E-05
p53 pathway	2.52	7.73E-03
Heterotrimeric G-protein signalling pathway	0.24	1.64E-02

Table 4-2. Overrepresented signalling pathways in the DE genes between BC and CP. The genes were mapped to PANTHERdb pathways and enrichment was calculated using the Fisher exact test and BH correction for multiple testing.

PANTHER Pathways	Fold-change	FDR
DNA replication	3.28	5.51E-03
De novo purine biosynthesis	2.82	2.70E-02
Apoptosis signalling pathway	2.03	3.04E-03
p53 pathway	1.91	4.85E-02

Table 4-3. TKI₆₀ genes are overrepresented in the lists of DE genes between AP and normal, BC and normal and BC and chronic phase.

Comparison	Total DE genes	TKI DE genes	p-value
AP vs Normal	2,810	48	<0.001
BC vs Normal	3,707	36	<0.001
BC vs CP	6,280	16	<0.001

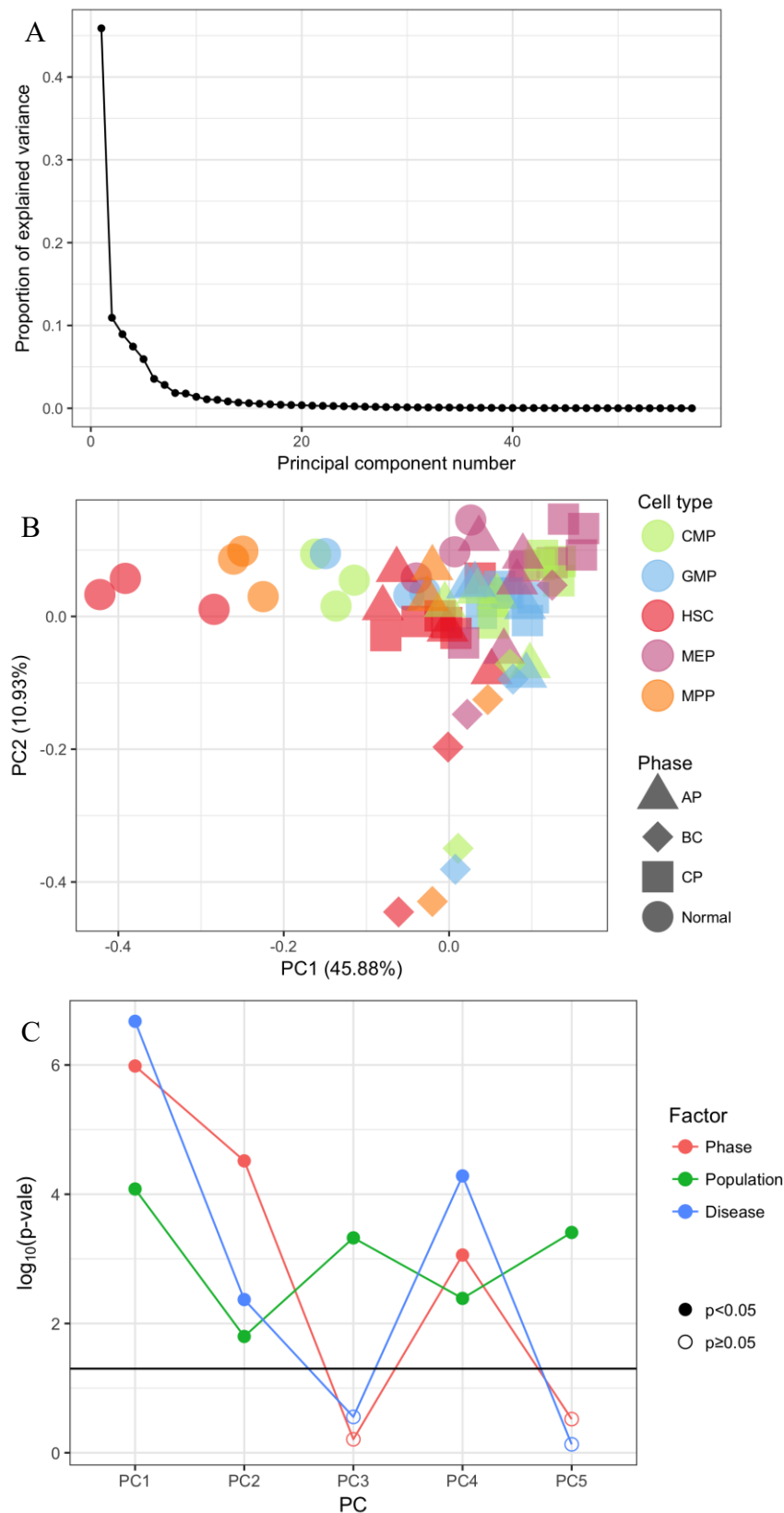


Figure 4-17. Selection for the TKIi₆₀ genes increases the variance explained by the principal component 1 and the association of disease phase with the principal components. (A) Percentage of variance explained by each PC. (B) The projection of the first 2 principal components allows discriminating the normal samples and the BC samples from the rest. (C) The PC1 was strongly associated with the presence of BCR-ABL1 and to a lower extent the phase of the disease and the cell population. The black horizontal line marks the threshold $p < 0.05$.

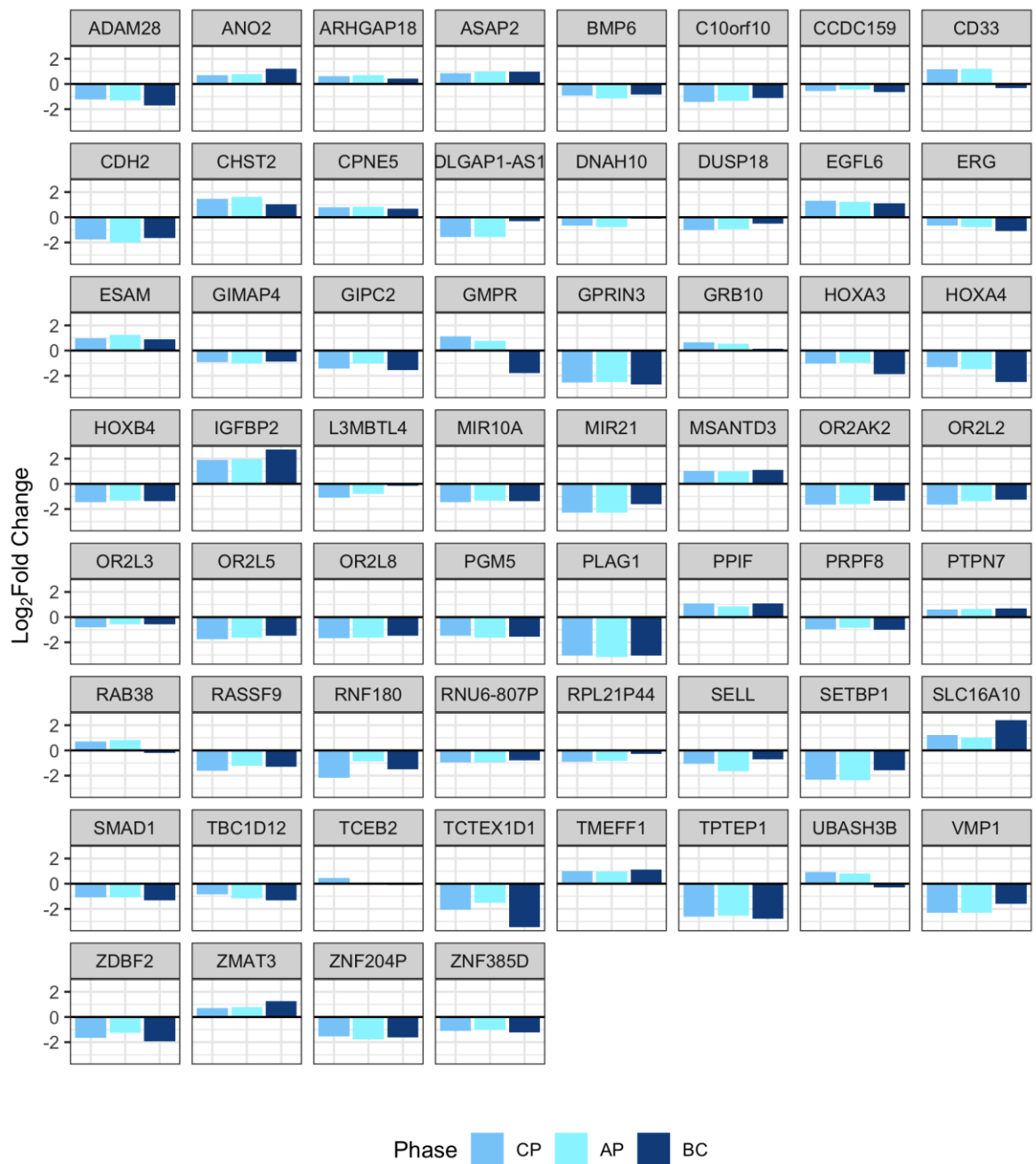


Figure 4-18. The direction of relative expression of the TKI_{i60} genes is constant in all phases of CML when compared with normal controls. Only *CD33*, *GMPT*, *RAB38* and *UBASH3B* had different direction of change in BC than in the other 2 phases. CP and AP had the same direction of change for all the genes.

4.5 Discussion

The TKI_i signature identified in Chapter 3 was found as the overlap between the differentially expressed genes in CML CP HSCs as compared with normal HSCs, and the genes not affected by TKI treatment. In order to assess whether this signature holds any potential translational value, transcriptional data describing various clinical scenarios was analysed, specifically (I) TKI response, (II) disease aggressiveness (classified retrospectively at diagnosis) and (III) disease progression. Additionally, as patients suffering from CML BC lack effective treatments, it was assessed if the direction of change on the genes in the TKI_i signature is constant among the 3 phases of the CML when compared to normal.

TKI treatment has proved to be very effective in managing CML, however not all patients respond to the treatment (Hughes et al., 2015) and the disease can evolve into accelerated phase or blast crisis. However, the current scores for predicting prognosis have only a limited sensitivity (16% for progression free survival using the EUTOS score (Hasford et al., 2011)). Therefore, the development of new molecular-based scores that may complement the existing ones could benefit those patients who are wrongly classified as low-risk.

Currently patients suffering from CML BC have a bad prognosis and most of them do not survive more than a year after diagnosis of the BC (Hehlmann, 2012). This is because the only therapeutic option for these patients is to undergo allogeneic haematopoietic stem cell transplant (aHSCT), which entails many complications and is not curative in all cases (Hehlmann, 2012, Baccarani et al., 2013, Saußele and Silver, 2015). With this in mind, it was decided to investigate if targeting the TKI₆₀ signature could constitute a new approach for the management of CML BC, in the hope of improving the survival of this group of patients.

The discovery of new biomarkers for predicting response to TKI is therefore of key importance in the management of CML patients. Using the RNRMW dataset it was assessed if the TKI₁₆₁ signature was indeed associated with TKI response. Despite the discrete difference between TKI responders and non-responders shown in the PCA (Figure 4-1, Figure 4-2) and the differential gene expression analysis (Figure 4-3) it was shown that the TKI₁₆₁ signature is overrepresented in the consistently differentially expressed probe sets (Figure 4-4) and that it is able to predict response to imatinib better than 97.1% of randomly generated probe sets of the same size. The classifier generated by (McWeeney et al., 2010) had a higher prediction value than the TKI₁₆₁ signature (Figure 4-11).

However, the McWeeney classifier was generated using the RNRMW dataset and optimised until the maximum AUC was achieved (e.g. by selecting the optimal number of probe sets). In contrast, the TKI₁₆₁ signature was discovered while answering a different biological question (i.e. existence of BCR-ABL1 TK independent genes in CML LSCs) and no optimisation was performed in the classifier (e.g. all the matching probe sets were used in the classifier). This suggests that the TKI₁₆₁ signature has a biological effect in CML LSCs by promoting survival, proliferation, self-renewal pathways and potentially LSC persistence. Thus, the ability to persist TKI treatment and the molecular mechanisms affecting the response to IM may be linked.

The PCA of AIAY dataset, which compares patients who suffered blast transformation within 3 years of diagnosis (aggressive disease) and patients who survived without suffering blast transformation for at least 7 years, revealed only minor association of the aggressiveness of the disease with the gene expression at diagnosis (Figure 4-12). PCA on TKI₁₆₁ matching probe sets showed even lower association of the aggressiveness of the disease with the gene expression at diagnosis. However, aggressiveness of the disease in the pre-TKI era was probably related to mechanisms different than the ones involved in TKI persistence. In contrast, PCA using the probe sets matching the McWeeney classifier increased the association of the aggressiveness of the disease with the gene expression at diagnosis. However, this association was confounded with an association of the age of the patient at diagnosis, which is also confounded in (McWeeney et al., 2010) as the median age of non-responders was 10 years higher than the median age of the responders' cohort (61-51). Older age is known to be a risk factor in CML and is included in the Sokal score model (Sokal et al., 1984), the first score used for predicting aggressiveness in CML. The Sokal score classification (including age) has also been associated with poorer response to imatinib (Nicolini et al., 2018, Breccia et al., 2018). The study of the effect of age in RNRMW dataset would have been important for understanding the gene expression changes between the groups. However, the lack of information in the age at diagnosis of each patient in the RNRMW datasets prevented further analysis.

A recent publication by the same group as RNRMW (Patel et al., 2018) showed no differences between nilotinib responders and non-responders. In this dataset the McWeeney classifier was unable to classify correctly the patients between responders and non-responders. Although it was not possible to test the ability of the TKI₁₆₁ signature to predict response to nilotinib in this thesis, the gene expression data will be requested for future work. Additionally, the age of the patients in the different groups of the study is not available, so it would not be possible to assess whether the lack of discriminative power of

the McWeeney classifier in this dataset is due to a more similar median age between the two groups.

Gene expression analysis on HSCs revealed that the TKI₆₀ signature is overrepresented in AP and BC compared with normal controls as well as in BC compared with CP. However, this should be validated using other datasets with independent samples as the CMLMC dataset was already used for the discovery of the TKI₆₀ signature and using the same data would provide limited insight. However, the results show that the TKI₆₀ signature's relative gene expression has consistent direction (up or down regulation) across all phases of the disease and it is able to discriminate normal HSCs from CML cells regardless of phase (Figure 4-17, Figure 4-18). There is also some evidence that the TKI₆₀ signature may discriminate BC HSCs from CP or AP HSCs, but this seems to be associated with principal component 2, which accounts for only 10% of the variance. Thus, this data suggests that the differential expression of the genes in the TKI₆₀ signature compared to normal HSCs is required in all three CML phases and therefore, targeting components of the TKI₆₀ signature may constitute a valid therapy in all the phases.

In summary, the TKI₁₆₁ signature seems to have some value as a potential biomarker for imatinib response in CML independent of age and its relative expression has a consistent direction of change compared to normal controls across all stages of CML. The development of novel biomarkers for TKI response could inform choice of treatment and monitoring in high-risk patients. Testing the predictive value of the TKI₁₆₁ signature in other datasets, including different TKIs, would confirm its value and its independence from confounding factors. Additionally, the development of therapeutic options targeting the components of the TKI₆₀ signature may constitute a new formula for the treatment of advance phases of CML.

5 Results (III): CD33 as a therapeutic target in CML CD34⁺ cells

5.1 Introduction

CD33 is a cell surface marker that has been used for identifying cells of the myeloid lineage and haematopoietic progenitor cells in the past decades (Andrews et al., 1989). Although CD33 has been associated with progenitor cells, it was thought that long-term repopulating cells did not have CD33 on their surface (Andrews et al., 1989). However, a recent publication by Connie Eaves (Knapp et al., 2018) has shown that CD33 is expressed in the human cord blood cells with longest repopulating potential.

CD33 is a CD33-related Siglec (sialic-acid-binding immunoglobulin-like lectin), a group of very similar proteins (9 in humans) mainly expressed in the innate immune system that are poorly conserved between species (Crocker et al., 2007). Most of them contain immunoreceptor tyrosine-based inhibitory motifs (ITIMs), making them inhibitory receptors by the recruitment of tyrosine and inositol phosphatases (Crocker et al., 2007). It has been shown that targeting CD33 with monoclonal antibodies has an important anti-proliferative effect in AML (Vitale et al., 2001) and a modest one in CML (Vitale et al., 1999). Additionally, it has been observed that CD33-related Siglecs function as endocytic receptors (Crocker et al., 2007), which allows targeting of CD33⁺ cells with ligand/antibody targeted chemotherapy (Crocker et al., 2007, Vitale et al., 2001).

This idea was applied by Pfizer in the development of the anti-leukaemic drug gemtuzumab-ozogamicin (GO, Mylotarg[®]), a conjugated antibody that binds CD33 and once internalised releases the cytotoxic molecule calicheamicin, which produces double strand breaks in the DNA that leads to the death of the cell (Naito et al., 2000). GO was commercialised between 2000 and 2010, when it was withdrawn from the market after a clinical trial revealed that GO was inducing hepatotoxicity in a number of patients (Jurcic, 2012). Since then, new trials have demonstrated the value of GO in the treatment of AML and it was re-approved by the FDA in September 2017 (Jen et al., 2018b). Although GO's therapeutic goal was to target AML cells, a publication showed that GO is effective at eliminating CML MNC (Herrmann et al., 2012).

The current chapter aims to assess the therapeutic potential of GO for the treatment of CML by studying its effects on CML CD34⁺ cells at the cellular and molecular levels. To do this the cells were treated with GO at different concentrations in the presence and absence of IM in different regimens: 72h GO±IM, 72h GO followed by 72h IM/no-treatment or 72h IM/no-treatment followed by 72h of GO. The effect of GO was measured

in cell proliferation, apoptosis, cell cycle phase, phosphorylation of H2AX, colony forming potential and global gene expression changes.

5.2 Results

5.2.1 *K562 cell line expresses CD33 and is sensitive to treatment with GO*

In order to test GO in our laboratory and to optimise different assays, it was decided to study the expression of CD33 in two different CML cell lines, K562 and BV173, and test the effect of GO on them. It was found that K562 cells express CD33 on their surface but BV173 do not (Figure 5-1A and B). Treatment of K562 cells with GO over 72h reduced the number of live cells to 50% of the NDC at $13.45 \pm 2.2 \text{ ng/mL}$ (Figure 5-1C) while it required $27.09 \pm 8.58 \text{ ng/mL}$ to achieve the same effect in BV173 cells (Figure 5-1D), as it was predicted by the detected higher CD33 expression on the cell surface of the latter. No reduction in cell numbers was observed at 24 or 48 hours (data not shown).

5.2.2 *Treatment with GO increases cell size in K562 cells*

Treatment with GO was observed to increase the cell size of K562 cells. The average viable cell size, as reported by the automatic EVE cell counter, increased $3.2 \mu\text{m}$ (from $12.6 \mu\text{m}$ to $15.8 \mu\text{m}$) when K562 cells were treated with GO 250 ng/mL for 72h when compared with the same cells grown for 72h with no drug (Figure 5-2A). The logarithmic transformation of this increase in size was observed to be correlated with the logarithm of the concentration of GO using Spearman's rank correlation test (Spearman's $\rho = 0.79$, $p = 1.411 \times 10^{-11}$).

5.2.3 *Treatment with GO has a modest impact on apoptosis in K562 cells*

The maximum effect of apoptosis observed on K562 cells treated with GO for 72h was a 27.4% reduction of viable cells (Annexin V DAPI) when compared with the NDC. The 50% of this maximum effect was observed at $9.5 \pm 2.4 \text{ ng/mL}$ (Figure 5-2B).

5.2.4 *GO and IM have additive effect when combined in K562 or BV173 cell lines*

Initial testing of the combination of GO with IM in K562 and BV173 cells revealed that it is able to target CML cells suggesting that GO is more effective on CD33-expressing cells, as was expected. However, TKIs have already demonstrated a high efficacy at managing CML and new treatments will potentially be used in combination with TKIs. Because of this it was decided to investigate the effect of the combination of GO and IM on the cell lines.

To investigate in more detail the interaction between the two drugs, K562 and BV173 cell lines were treated with increasing concentrations of GO (0 to 500 ng/mL for one plate of K562 cells and 0 to 50 ng/mL for the other K562 plate and BV173) and IM (0 to $5 \mu\text{M}$; Figure 5-3) and the effect analysed using resazurin, a dye that changes colour based on

metabolic activity. Using the Bliss' method (Bliss, 1939) it is possible to study if two different drugs' combined effect differs from the one expected effect of the combination if both drugs act independently. Both drugs were found to have mostly additive effects although they worked in a slightly antagonistic manner, especially at higher concentrations of GO. The average Bliss index, which equals "1 - % viable" compared with the NDC, was -1.859 for BV173 (Figure 5-4A) and -0.812 and -2.18 for low and high concentrations of GO in K562 cells, respectively (Figure 5-4B and C). Although this result might not be the expected one, the lack of effect observed at high concentrations of GO might be caused by the increase in size observed in the cells after treating them with GO, which could increase the metabolic rate of the surviving cells compared with the NDC (an increase of 3.2µm in diameter means an increase of 17.16 µm³ in volume, approximately). Therefore, this could reduce the observed response to GO when using resazurin than when using cell counts.

5.2.5 Binding of GO to the cell surface is dose dependent

The binding of GO to the cell surface was measured by flow cytometry after one hour of treatment using an anti-human IgG antibody, which binds to GO, a humanised IgG. It was shown that the amount of GO bound to K562 cells increased in a dose dependent manner until 62.5ng/mL, when it potentially saturated all epitopes (Figure 5-4D). At later time points it was not possible to detect any GO on the cell surface or in the inside of the cell. This can potentially be explained by the internalisation of CD33 after binding of GO and its degradation in the lysosome.

5.2.6 GO induces DNA damage in a dose dependent manner and it is independent of IM treatment

The proposed mechanism of GO that calicheamicin produces double strand breaks in the DNA, which would be followed by the cell activating cell death pathways (Naito et al., 2000). Histone H2AX is known to participate in DNA double strand breaks and its phosphorylation is a requirement for the recruitment of some of the factors involved in DNA double strand break repair (Rogakou et al., 1998, Paull et al., 2000). Thus, an increase in H2AX phosphorylation on serine 139, or γH2AX, would suggest an increase in the formation of double strand breaks in the DNA and, therefore, detecting this increase would inform about the double strand breaks. During this project, the presence of γH2AX was measured by flow cytometry in permeabilised cells. It was possible to observe an increase in the presence of γH2AX with increased concentrations of GO (Figure 5-5A) while it was observed that IM neither increased the presence of γH2AX alone (Figure 5-5B) nor in the presence of GO (Figure 5-5C) in K562 cells.

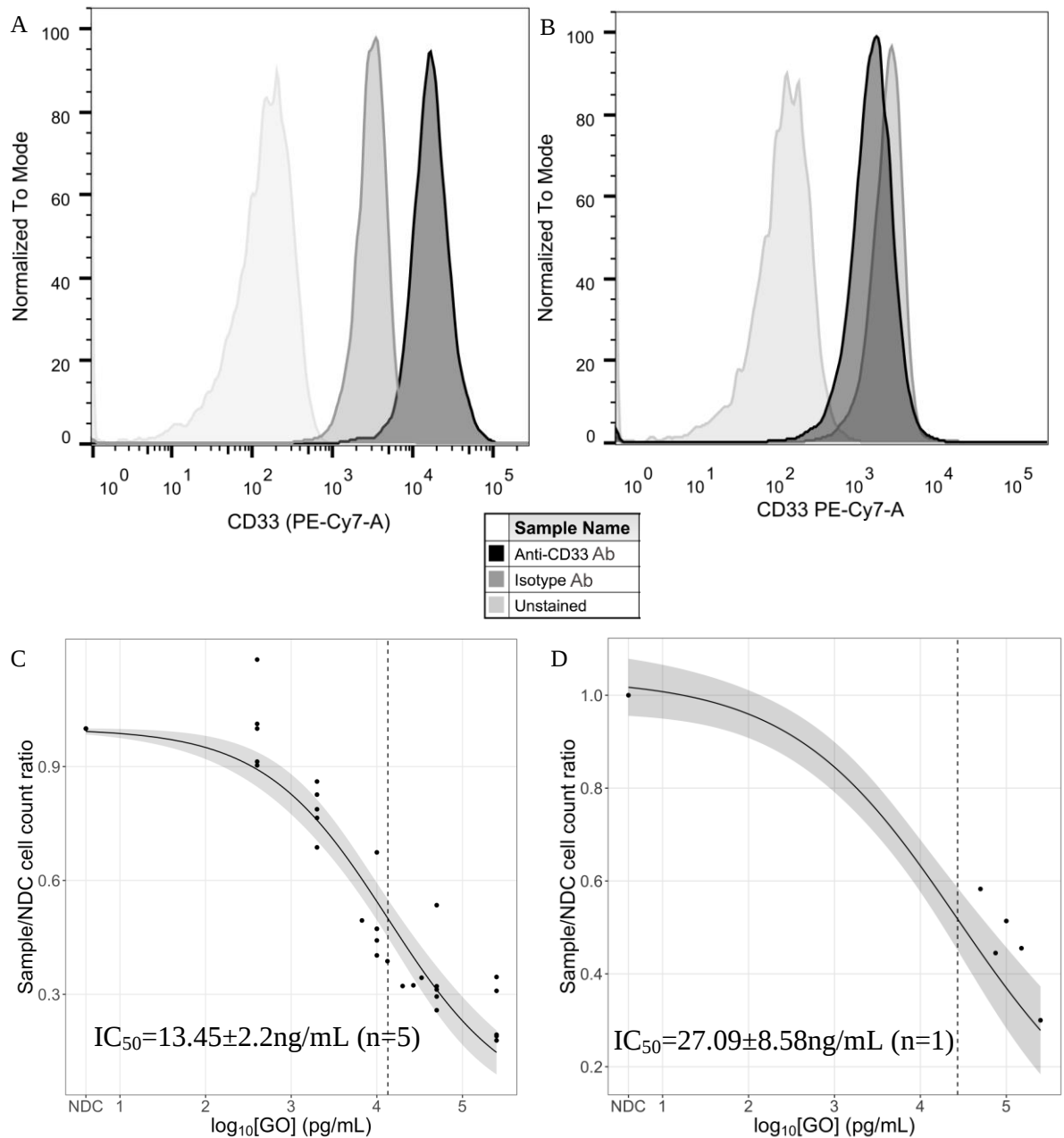


Figure 5-1. K562, a CD33-expressing cell line, has a lower IC_{50} than BV173, a non CD33-expressing cell line. (A) K562 cell line presents higher fluorescence intensity values when the cells are incubated with an anti-CD33 antibody than when incubated with an isotype control. (B) BV173 cell line shows no increase in fluorescence intensity between the anti-CD33 and the isotypes antibodies, suggesting no expression of CD33 on the cell surface. Concentration-response curves in response to difference concentrations of GO in (C) K562 and (D) BV173.

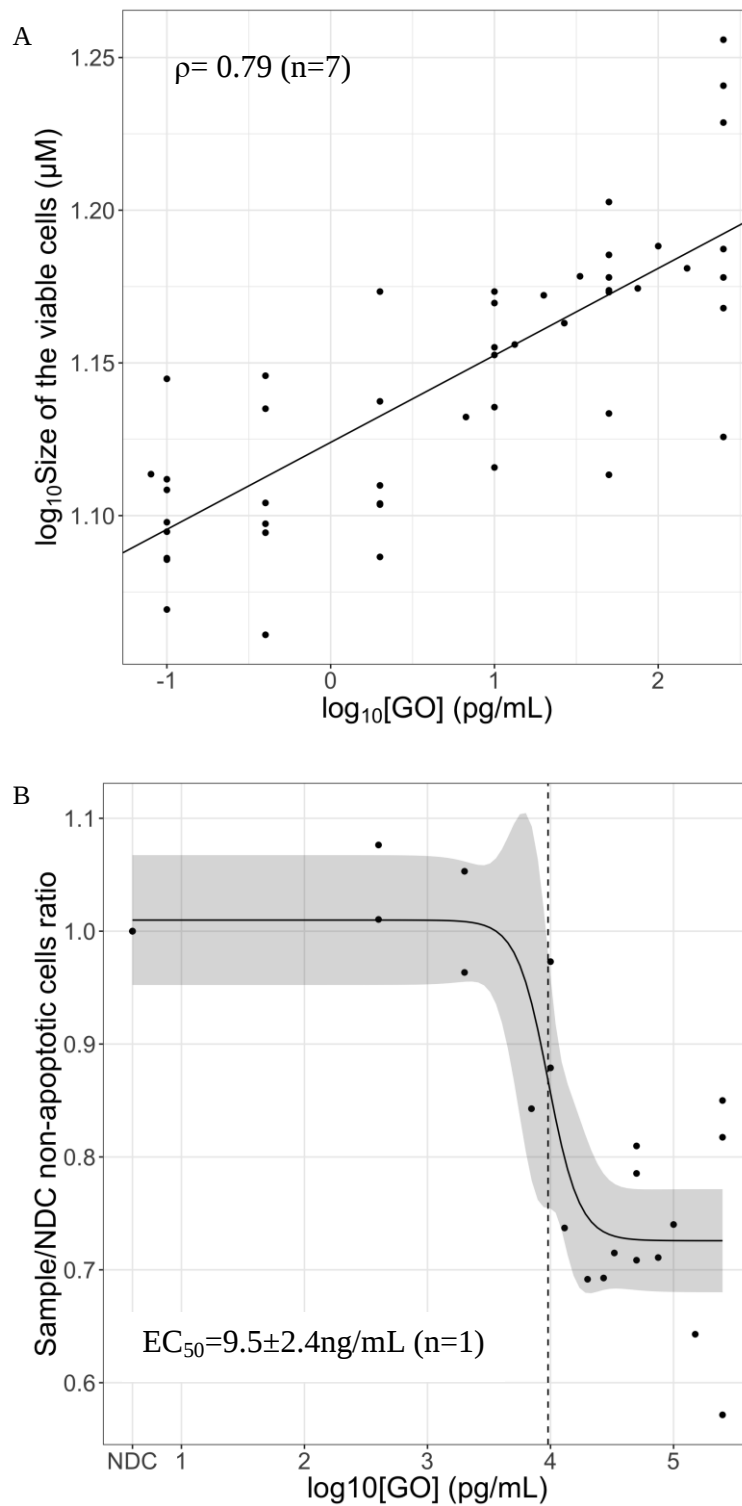


Figure 5-2. GO increases size of K562 cell line and induces apoptosis. (A) The average size of the viable cells increases as the concentration of GO increases when treating K562 cell line. (B) Decrease of viable K562 cells (cells negative for Annexin V and DAPI) when treated with increasing concentrations of GO.

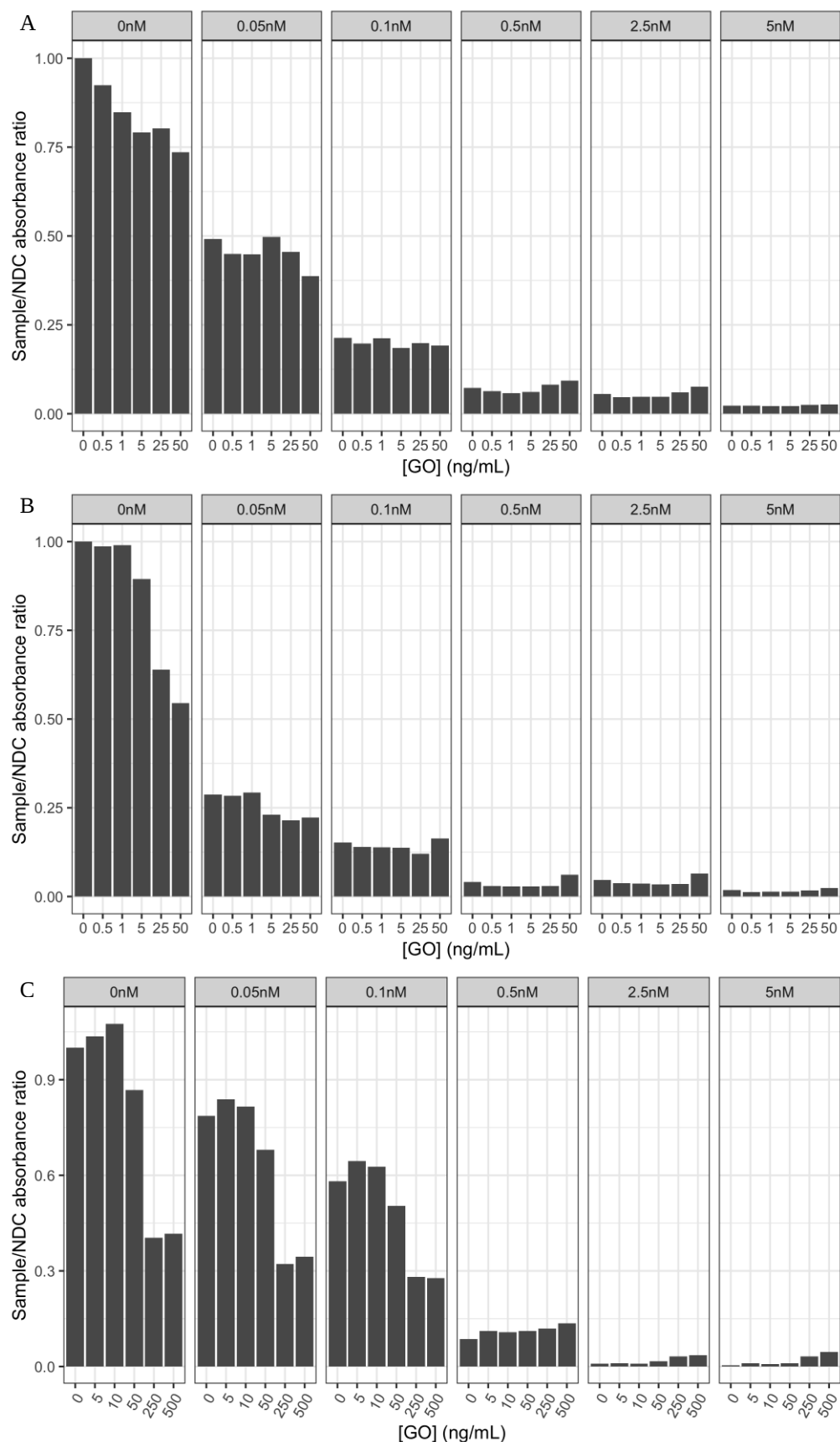


Figure 5-3. Relative cell concentrations to NDC based on resazurin absorbance. IM concentration is stated in the grey box on top of the columns. Each value corresponds to a single observation. (A) Relative cell concentration of BV173 cells between 0 and 50ng/mL of GO. (B) Relative cell concentration of K562 cells between 0 and 50ng/mL of GO. (C) Relative cell concentration of K562 cells between 0 and 500ng/mL of GO.

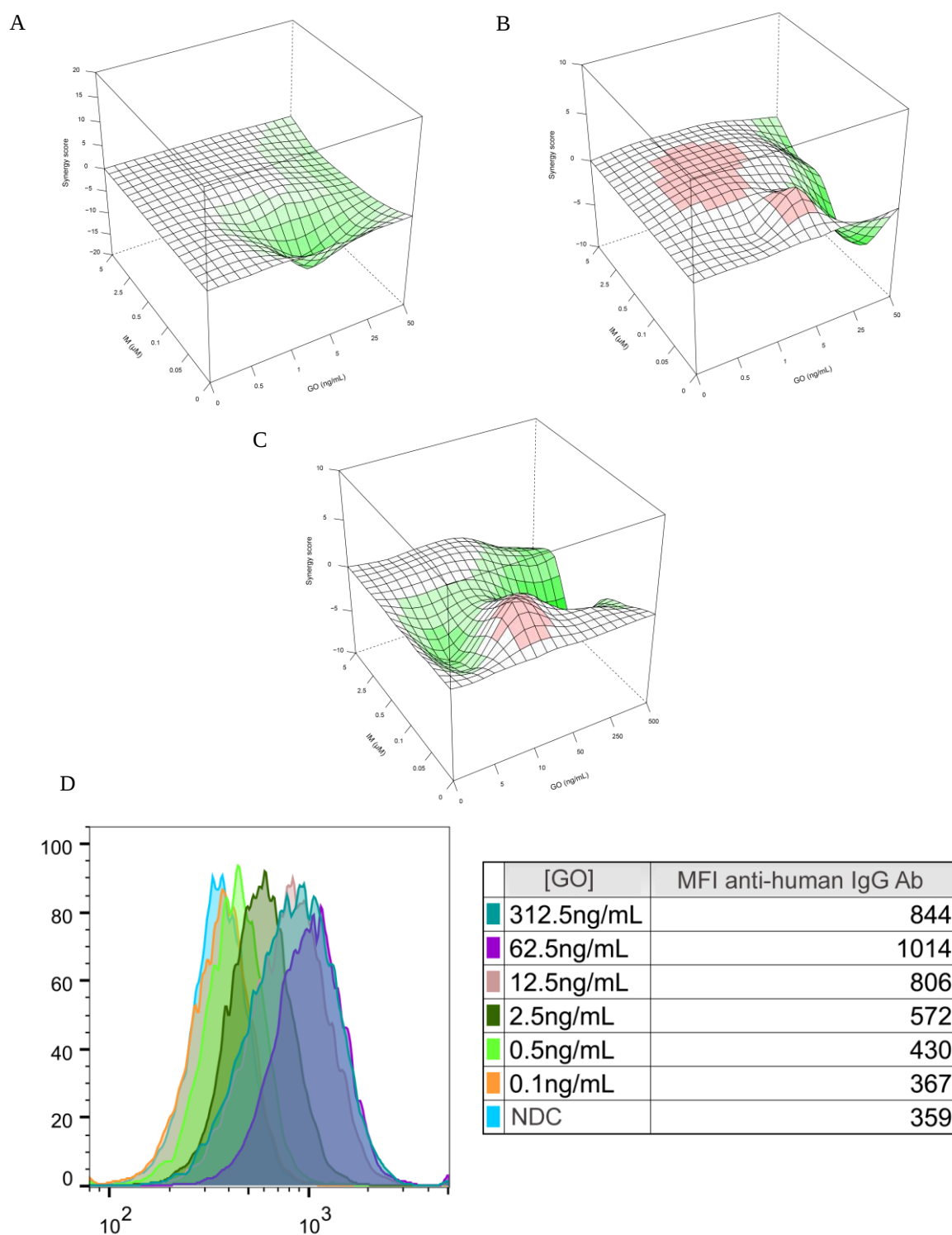


Figure 5-4. Bliss coefficients for each IM-GO combination and binding of GO to the cell surface. Green represents antagonism and red synergism. (A) Coefficients for BV173 cells. (B) Coefficients for K562 cells treated with 0-50ng/mL of GO. (C) Coefficients for K562 cells treated with 0-500ng/mL. (D) It is possible to observe, both in the density plots and the table that the amount of GO that bind to the cells increases with GO concentration until approximately 62.5ng/mL, when it reaches its peak. Data from a representative experiment.

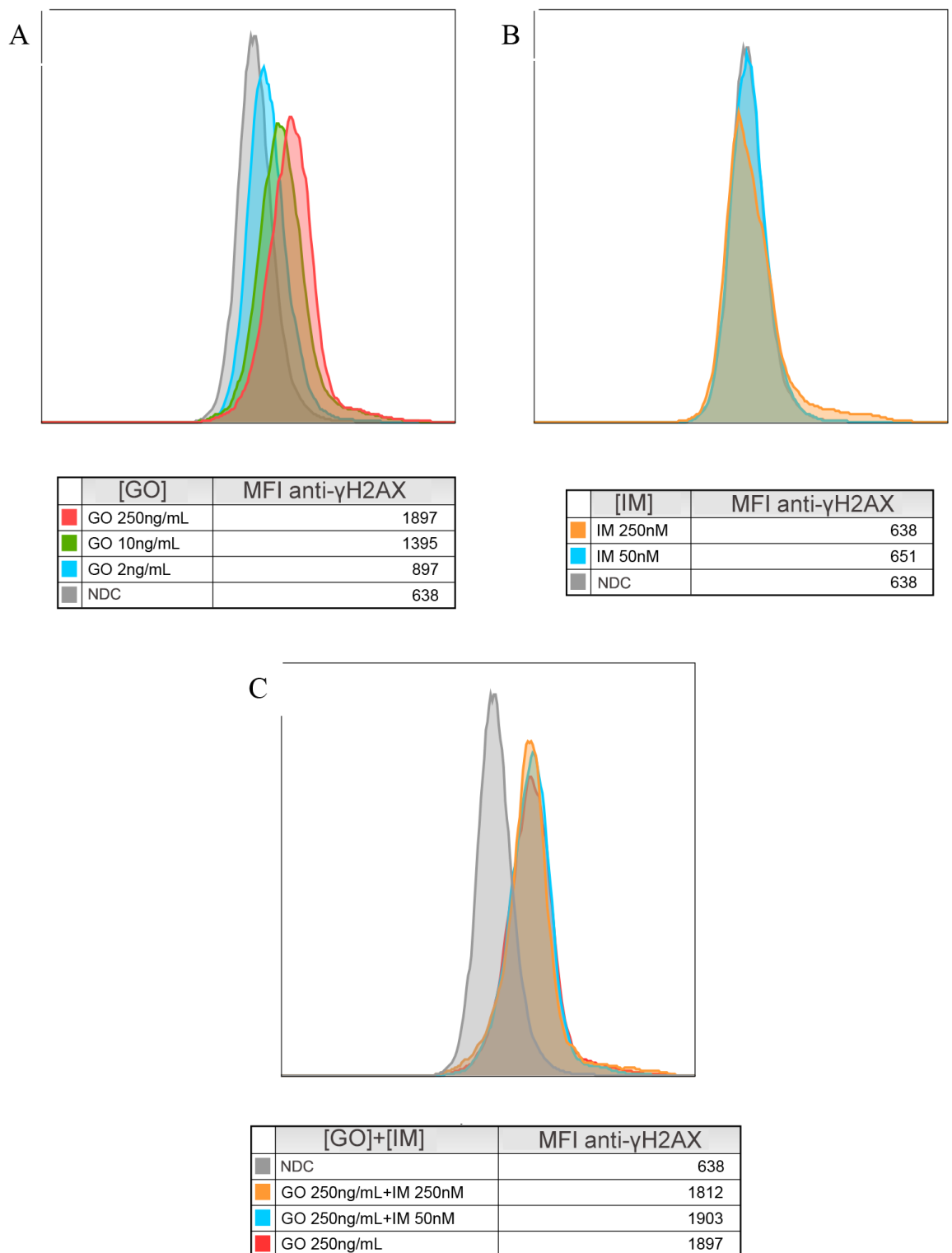


Figure 5-5. The level of γ H2AX increases in response to GO but it is not affected by IM in K562 cells. (A) The levels of γ H2AX increase with increasing concentrations of GO. (B) The level of γ H2AX is not affected by the concentration of IM compared with the NDC in K562. (C) The combination of GO with different concentrations of IM does not increase the level of γ H2AX compared with a sample treated with the same concentrations of GO. Figure shows the results of a representative experiment.

5.2.7 CML CD34⁺ cells express CD33 on the cell surface at similar levels to nCML CD34⁺ cells

Expression of CD33 on the cell surface was measured before the start of treatment on both CML and nCML CD34⁺ cells. In order to normalise the intensity values between patients, the unstained control's median fluorescent intensity (MFI) was subtracted from both the test measurement and the isotype control MFIs. Finally a ratio between the test and the isotype control was calculated. This allowed controlling for both background noise and non-specific binding of the antibody. CML cells presented a more heterogeneous expression of CD33 than nCML cells but the medians were very similar, which suggested that both groups of patients have similar levels of CD33 on their cell surface ($p=0.53$) despite the differences found at the transcript level in Chapter 3. Interestingly, and despite the published reports to the contrary (Herrmann et al., 2012), the single blast crisis sample that was analysed was found to have higher levels of CD33 on the cell surface than chronic phase cells (Figure 5-6A).

5.2.8 Binding of GO to the cell is dependent on the levels of CD33 on the cell surface

The levels of GO bound to the cells were measured after 1 hour of treatment at 100ng/mL. The normalised MFI of the anti-human IgG antibody ($MFI_{\text{test}}/MFI_{\text{unstained}}$) was compared with the normalised CD33 MFI for each patient's cells. The levels of GO bound to the cell surface were found to be highly correlated with the levels of CD33 of the surface of the cells (Spearman's $\rho=1$, $p=0.017$; Figure 5-6B).

5.2.9 IM treatment does not affect CD33 levels on the surface of the cells

While was confirmed that the mRNA levels of *CD33* do not change after TKI treatment (Figure 3-11C), GO interacts directly with the protein expressed on the cell surface and, therefore, a constant expression of the protein is required for a consistent GO effect. The amount of CD33 expressed on the surface of the cells was measured using flow cytometry with treated with IM for 72h and untreated cells. No differences were found in the expression of CD33 on the cell surface either in CML or in nCML (Figure 5-6C).

5.2.10 GO reduces cell number in CML CD34⁺ cells while having a minor effect on nCML CD34⁺ cells after 72h of treatment

Once the effect of GO had been confirmed in K562 cells (Figure 5-1, Figure 5-5), and its basic mechanism of action confirmed, it was decided to test its effect on CML CD34⁺ cells. Patient samples that did not respond to IM in the clinic were favoured for inclusion in the

experiment as this population is the one that potentially could benefit from GO the most. Patient information is summarised in (Table 2-1, Table 2-2).

Initially it was decided the range of GO concentrations to be tested should be from 0 to 3000ng/mL as this is the known peak plasma concentration for the first infusion in the treatment regimen that was approved for AML until 2008 (EMA, 2008), which consisted of two infusions of 9mg/m² of GO in a period of 14 days. This regimen had a high risk of causing liver problems and veno-occlusive disease in patients receiving haemopoietic stem cell transplant (EMA, 2008). However, as the effect at 1000ng/mL in CD34⁺ cells, which approximates the peak plasma concentration for the recently approved 3mg/m² regimen for patients with AML, was already close to maximum effect it was decided to study lower concentrations of the drug. IM when used was added at 2μM, which approximates the plasma mean concentration (Druker et al., 2001b), and the cells were cultured in the presence of PGF.

It was shown that treatment with GO for 72h was very effective at eradicating CML CD34⁺ cells. The IC₅₀ was found to be 136.29±22.86ng/mL (Figure 5-7), a concentration that is more than 7 times lower than the peak plasma concentration with the treatment regimen that is currently approved for the treatment of AML. The differences in the sensitivity to IM could explain heterogeneity to the results of the combination treatment. This means that although the IC₅₀ for the combination (using the IM only sample as reference instead of the NDC) was similar to the single treatment at 195.42ng/mL (p=0.067; Figure 5-7), the standard error increased to 108.27ng/mL. The similarity between the two IC₅₀s suggested again that both drugs' effects are independent of each other. In order to validate this hypothesis, the combination effect of each patient was compared with the expected effect of the combination based on its response to IM alone and each concentration of GO using Bliss equation (Bliss, 1939). The observed effect was in average 1.56% lower than the expected effect for the interaction and was found different than 0 (p=0.04213, t=-2.95, DF=4). Although this suggests an antagonistic effect, the magnitude of the effect is small and the overall effect is close to that obtained from an additive effect.

The nCML controls were treated under the same conditions than the CML cells, presenting an IC₅₀ of 2,643.4±1,756.1ng/mL (Figure 5-7), which is more than 19 bigger than for CML CD34⁺ cells (p=1.75x10⁻⁴²). The large standard error could be explained by the reduced number of data points at that end of the curve. Most samples were treated with a maximum concentration of 1000ng/mL (as mentioned before) and, therefore, the reduced number of

data points increases the confidence interval at that range of the curve. Although IM does not seem to have a noticeable effect on nCML cells at low concentrations of GO, it seems to potentiate the effect of GO at concentrations higher than 100ng/mL. The combination treatment presented an IC_{50} of 892.23 ± 470.58 ng/mL using the IM-only samples as minimum effect (Figure 5-7), being significantly different from the IC_{50} of the single treatment ($p=0.0451$) and the IC_{50} of the combination treatment on CML ($p=4.39 \times 10^{-5}$). Although Bliss independence test found the combination treatment to be 4.28% more effective than the expected effect, this was not found to be statistically significant ($p=0.5918$, $DF=6$), suggesting that GO and IM have additive effects on nCML CD34⁺ cells.

5.2.11 GO induces apoptosis in CML CD34⁺ cells after 72h of treatment

Viability was also measured as the percentage of cells that were not stained with AnnexinV or DAPI, markers for apoptosis and cell death respectively. In order to facilitate the comparison between the different groups, the absolute percentage of AnnexinV⁻ DAPI⁻ cells was normalised with the absolute percentage of the NDC for each patient. IM had an important effect on CML cells, reducing the percentage of viable cells to 48% of the NDC while having no effect on nCML (93.75% of the NDC).

The effect of GO in this assay was equally effective in both CML and nCML, which IC_{50} s were 130.95 ± 47.22 ng/mL and 81.1 ± 27.8 ng/mL respectively ($p=0.44$; Figure 5-8). There was no statistically significant difference between the two combination arms (CML 52.4 ± 32.95 ng/mL; nCML 26.45 ± 8.9 ng/mL; $p=0.49$; Figure 5-8) or between the single arms and the combination treatments ($p_{CML}=0.41$; $p_{nCML}=0.16$).

This data suggests that GO is able to target those cells that are insensitive to IM, as the IC_{50} is not significantly different between the single treatment (GO) and the combination in CML (i.e. the same proportion of CML cells are killed by GO independently of the presence of IM). The discrepancy between the cell counts and the apoptosis staining for the difference of the CML and nCML IC_{50} s may need to be addressed with additional assays. This could be caused by the survival advantage that BCR-ABL1 confers on the viable cells. Another explanation could be that GO reduces the number of viable CML cells by other mechanism in addition to apoptosis, such as cell cycle arrest or other cell death mechanism.

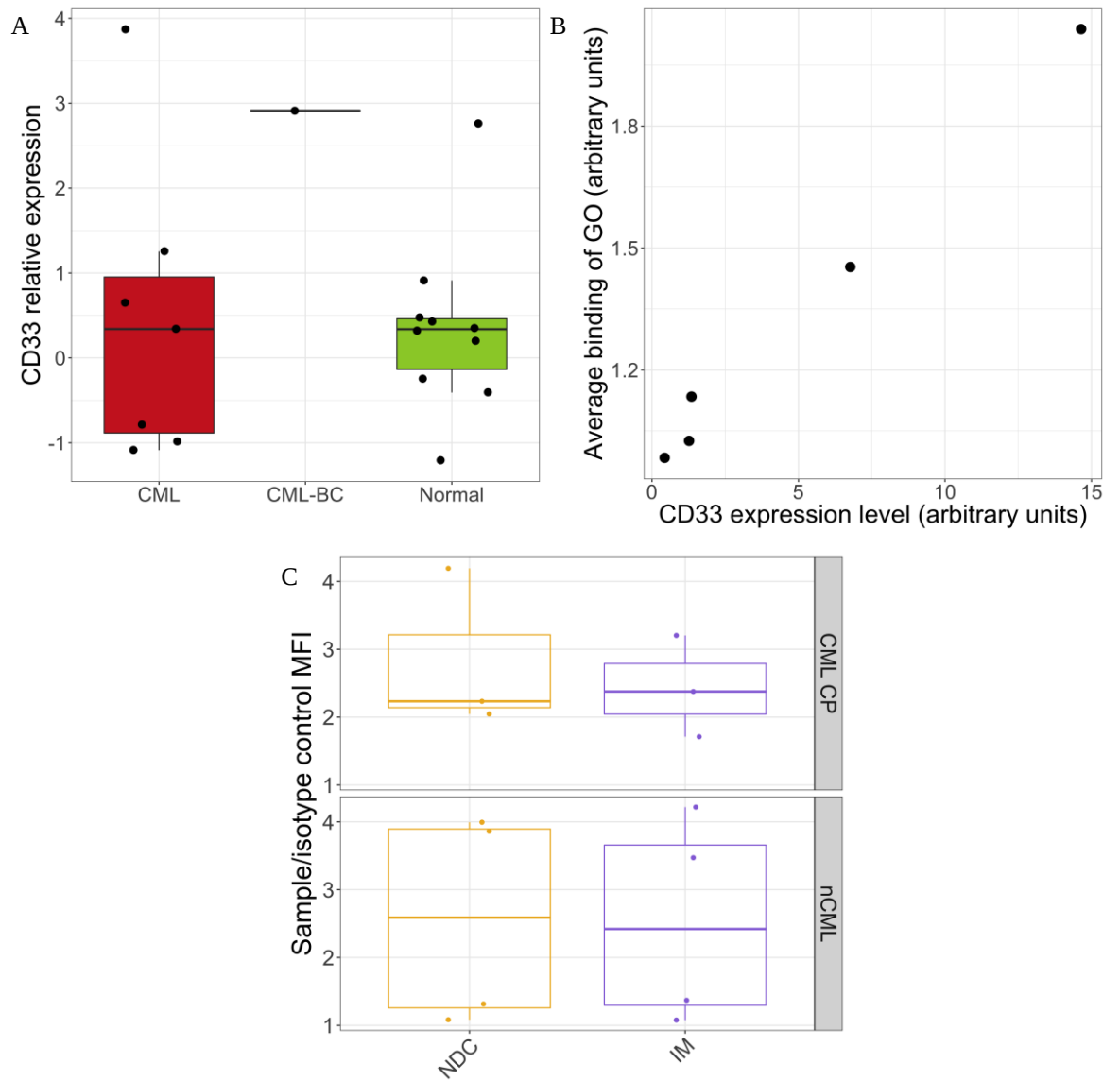


Figure 5-6. Expression levels of CD33 on the cell surface were the same between CML and nCML and are not affected by IM. (A) The levels of CD33 on the cell surface were calculated as relative expression compared with the isotype control before starting treatment with GO (baseline) and no differences were found between CML and nCML. One blast crisis CML samples was analysed, which presented high levels of CD33 on the cell surface. (B) The levels of CD33 on the cell surface were associated with a higher amount of GO (human IgG) bound to the cell surface. (C) Treatment with IM for 72h did not affect the expression level of CD33 on the cell surface compared with the NDC. Individual dots represent individual observations.

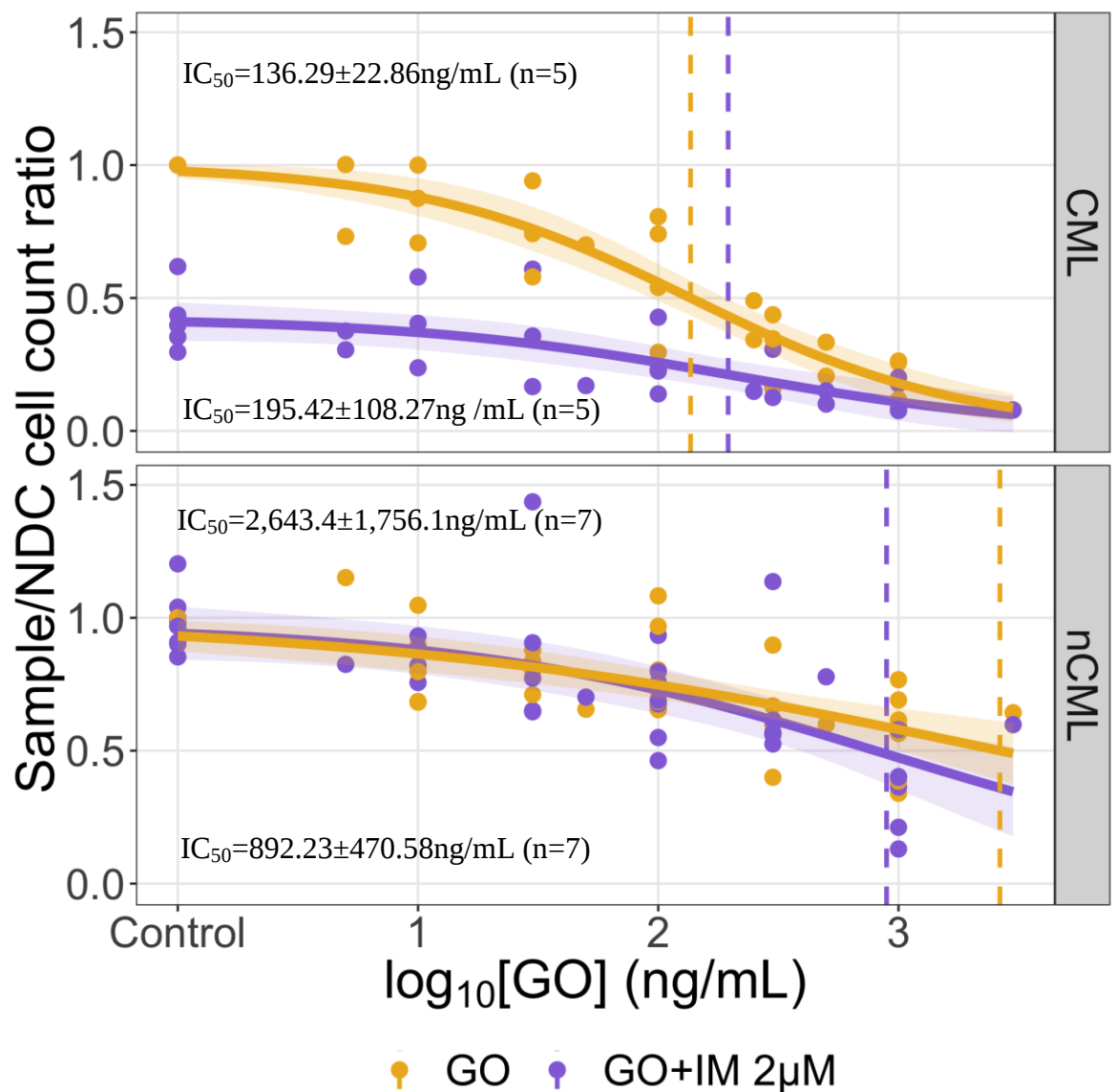


Figure 5-7. Treatment with GO reduces CML cell number after 72h of treatment. GO treatment was able to reduce the number of CML cells at lower concentrations than nCML cells, as can be seen in the two different IC₅₀s. Also, GO further reduced the number of CML cells when combined with IM while having a limited effect on nCML cells at concentrations lower than 300ng/mL. The IC₅₀s at the top of each panel correspond to the GO single treatment while the IC₅₀s at the bottom correspond to the combination treatment. Response to the treatment is represented with the solid line with 95% confidence interval (shadowed area). The dashed lines represent the IC₅₀s. Individual dots represent individual observations.

5.2.12 GO eliminates colony forming cells (CFCs) in both CML and nCML after 72h of treatment

In order to study the effect of GO in the progenitor population contained in the CD34⁺ cells, 3,000 viable cells (trypan blue negative) were plated in MethoCult™ H4034. Colonies were left to grow for approximately 15 days before counting and scoring them as erythroid (E), granulocyte (G), macrophage (M), granulocyte-macrophage (GM) or granulocyte-erythrocyte-macrophage-megakaryocyte (GEMM) type colonies.

Although the number of colonies for every 3,000 viable cells seemed to be higher in the nCML cells (Figure 5-9A), this difference was not statistically significant, potentially because of lack of power. It is also worth noticing that IM alone did not reduce the colony count neither in CML nor in nCML. IM did not show any reduction in cell number or viability in nCML and it has been reported to increase the expression of HSC-related genes (Charaf et al., 2016), so a reduction of CFCs was not expected. In CML, IM has an antiproliferative effect (Graham et al., 2002), so a change in the concentration of CFCs was also not expected. In CML the combination treatment with 100ng/mL of GO had a significantly lower number of colonies than both the NDC ($p=0.017$) and IM alone ($p=0.027$). In nCML, the NDC presented significantly more total number of colonies than 100ng/mL of GO alone ($p=0.033$) and the combination treatment with 30ng/mL ($p=0.03$) or 100ng/mL of GO ($p=0.013$), while IM alone had significantly more total colonies than the combination treatment with 30ng/mL ($p=0.047$) or 100ng/mL of GO ($p=0.021$). No differences between CML and nCML were found for any of the treatments either in the total colony count (Figure 5-9A) or in the different subtypes (Figure 5-9B). For individual types of colonies CML had significant differences in E ($p=0.025$) and M colonies ($p=0.014$; Figure 5-9B upper panel) while nCML samples presented significant differences only in E colonies ($p=0.013$; Figure 5-9B lower panel).

The total concentration of CFCs in the original culture (before culturing the cells in MethoCult™) was calculated by multiplying the number of CFCs per cell by the number of cells per mL. It was possible to observe an increase in the number of total CFCs in the CML NDC when compared with nCML ($p=0.008$) but not for the treated cells (Figure 5-10A). The same could be observed for the M colonies ($p=0.0002$) but no significant differences were observed for any of the other colony subtypes (Figure 5-10B). The total CFC concentration in CML was significantly reduced after the combination treatment with 100ng/mL of GO ($p=0.037$) while in nCML all GO treated arms had a significantly lower concentration of CFCs than the NDC ($p_{\text{single}30} = 0.029$; $p_{\text{single}100} < 0.001$; $p_{\text{combination}30} < 0.001$; $p_{\text{combination}100} < 0.001$) and all but 30ng/mL of GO alone when compared with IM

alone ($p_{\text{single100}} < 0.001$; $p_{\text{combination30}} = 0.0014$; $p_{\text{combination100}} < 0.001$; Figure 5-10A). The nCML samples presented more significant differences even though the magnitude of the effect was smaller than in CML, as can be observed in Figure 5-10A. This could be caused by the higher variability on cell proliferation observed in the NDC in CML samples (in the original cultures, before transferring the cells into MethoCult™). For the individual subtypes, CML presented significant differences in E ($p=0.006$) and G ($p=0.031$) CFCs while nCML had significant differences in the E ($p<0.001$), M ($p<0.001$) and GM ($p=0.003$) subtypes (Figure 5-10B).

5.2.13 Detection of LTC-IC after 72h of GO treatment is patient dependent and no differences were found between treatments

Although the standard CFC assay allows determination of the amount of progenitor cells present in a given population it is not appropriate for determining the number of stem cells in the same cell population. In order to determine the effect that GO had on the stem cells, after the treatment the cells were cultured in the presence of support stroma cells for 6 weeks and plated on MethoCult™ H4034 at the end of that period for determining how many CFCs remained. This assay works on the assumption that stem cells, having self-renewal capacity, are able to maintain themselves over the 6 weeks while progenitor cells will exhaust and differentiate in the same period of time.

LTC-ICs are a challenging assay with a high variability between patients. This prevented to finish the experiment for some of the conditions, as some conditions got fungal contamination over the 6 weeks. This is the reason why the amount of observations for between conditions differs in Figure 5-11 and Figure 5-12. Although the number of colonies counted in nCML patients was higher than in CML, both as per cell (Figure 5-11) or corrected by the cell concentration in the original culture (Figure 5-12), which is consistent with the literature and the proliferative phenotype conferred by BCR-ABL (Charaf et al., 2016), no significant differences were found between CML and nCML. No differences were found either between the different treatments in the total counts or in any of the colony subtypes. On the other hand, there was a significant difference between samples of the same condition ($p_{\text{CML}}<0.001$ and $p_{\text{nCML}}=0.002$), which shows the high heterogeneity of the LTC-IC data within the same group.

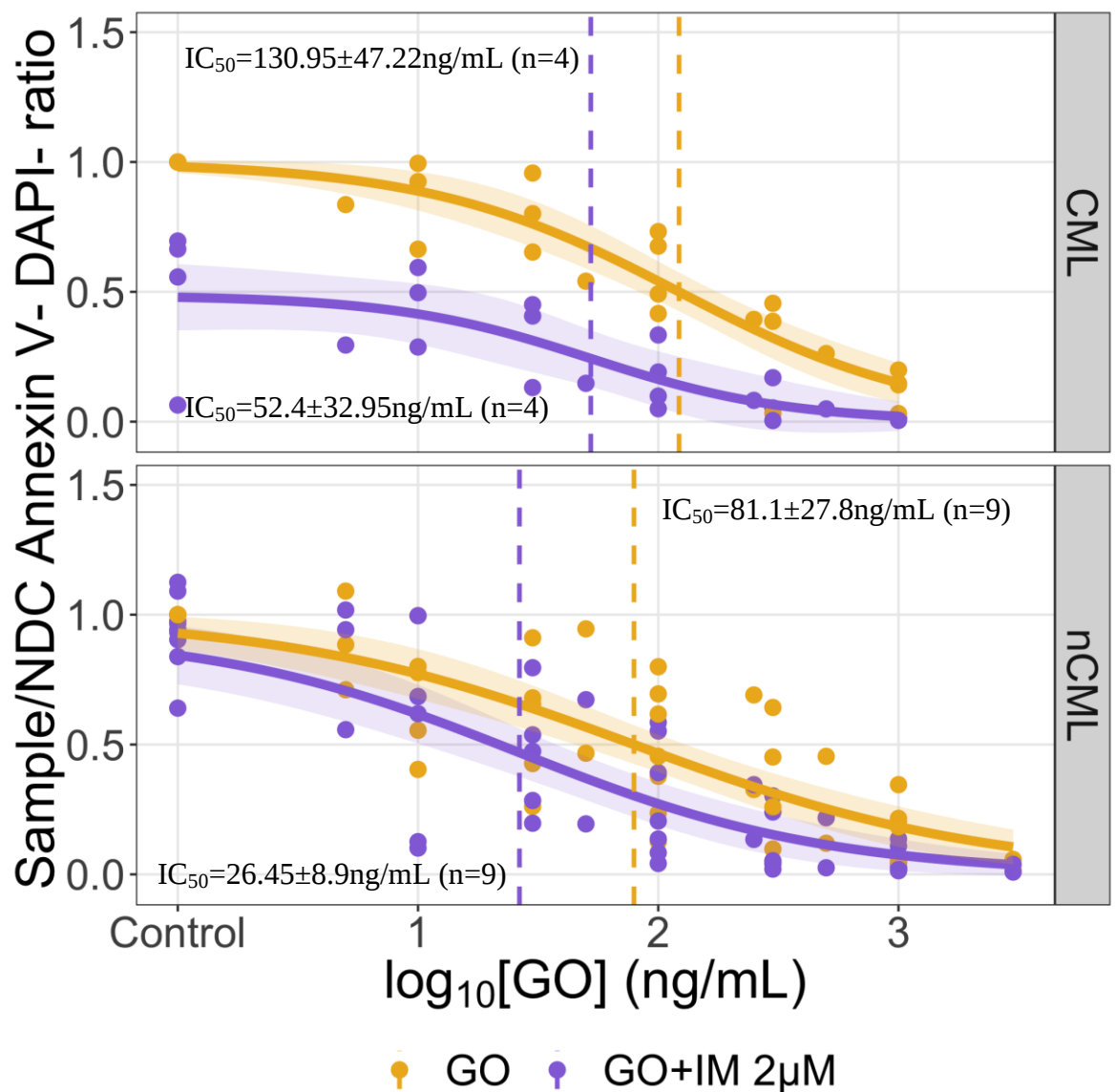


Figure 5-8. GO reduced the percentage of viable cells in both CML and nCML. The percentage of viable cells (Annexin V⁺ DAPI⁺) was reduced similarly in CML and nCML as it was between GO single treatment and its combination with IM. The IC₅₀s at the top of each panel correspond to the GO single treatment while the IC₅₀s at the bottom correspond to the combination treatment. Response to the treatment is represented with the solid line with 95% confidence interval (shadowed area). The dashed lines represent the IC₅₀s. Individual dots represent individual observations.

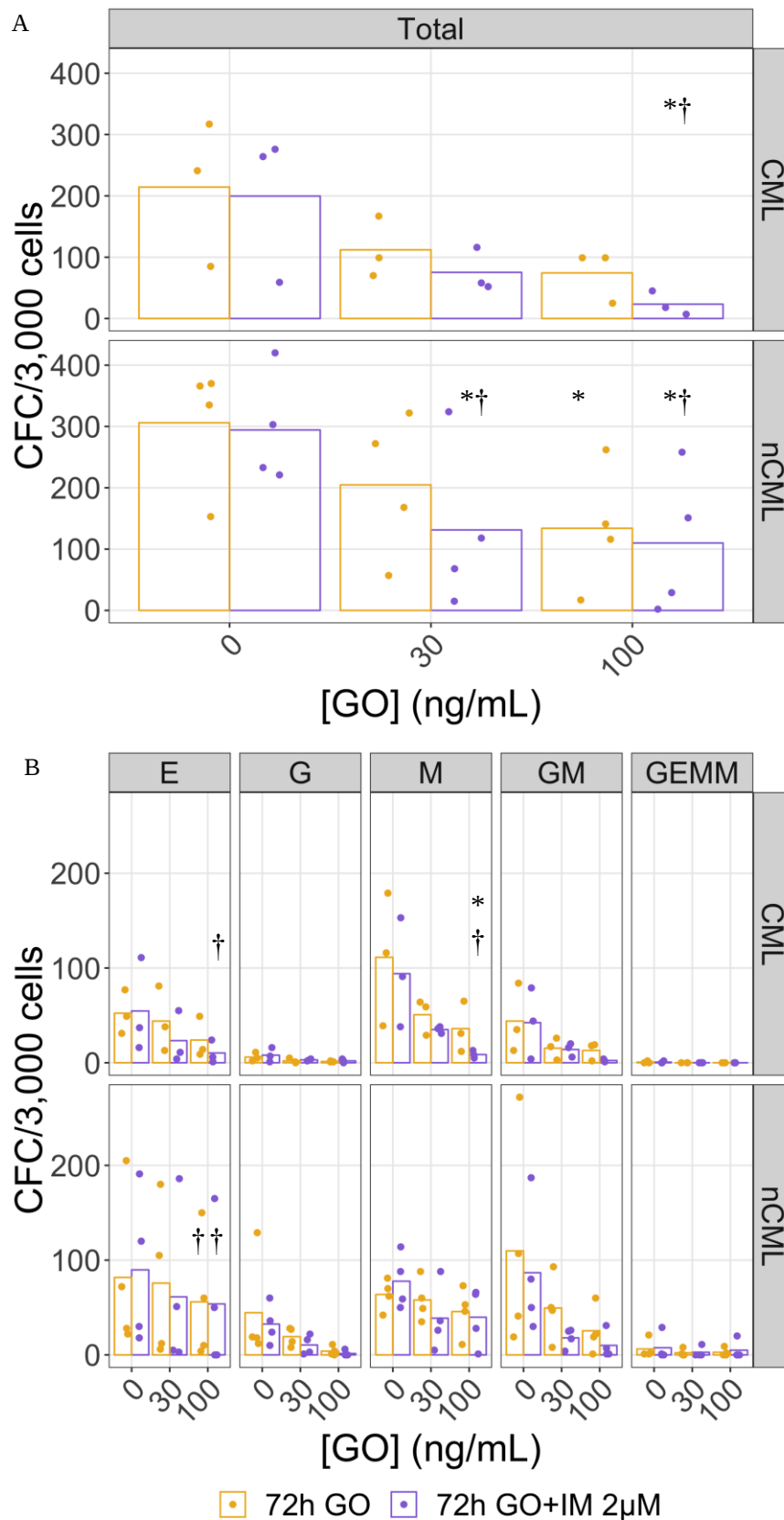


Figure 5-9. Colony counts after 72h of treatment with GO±IM. (A) Total colony counts for each treatment and condition. IM has a limited effect on its own but it further reduces the colony counts in CML when it is combined with GO. (B) Colony counts for each subtype. Individual dots are the colony counts for each patient and bars are the mean. Significant differences with the NDC are represented with * and significant differences with IM only are represented with †.

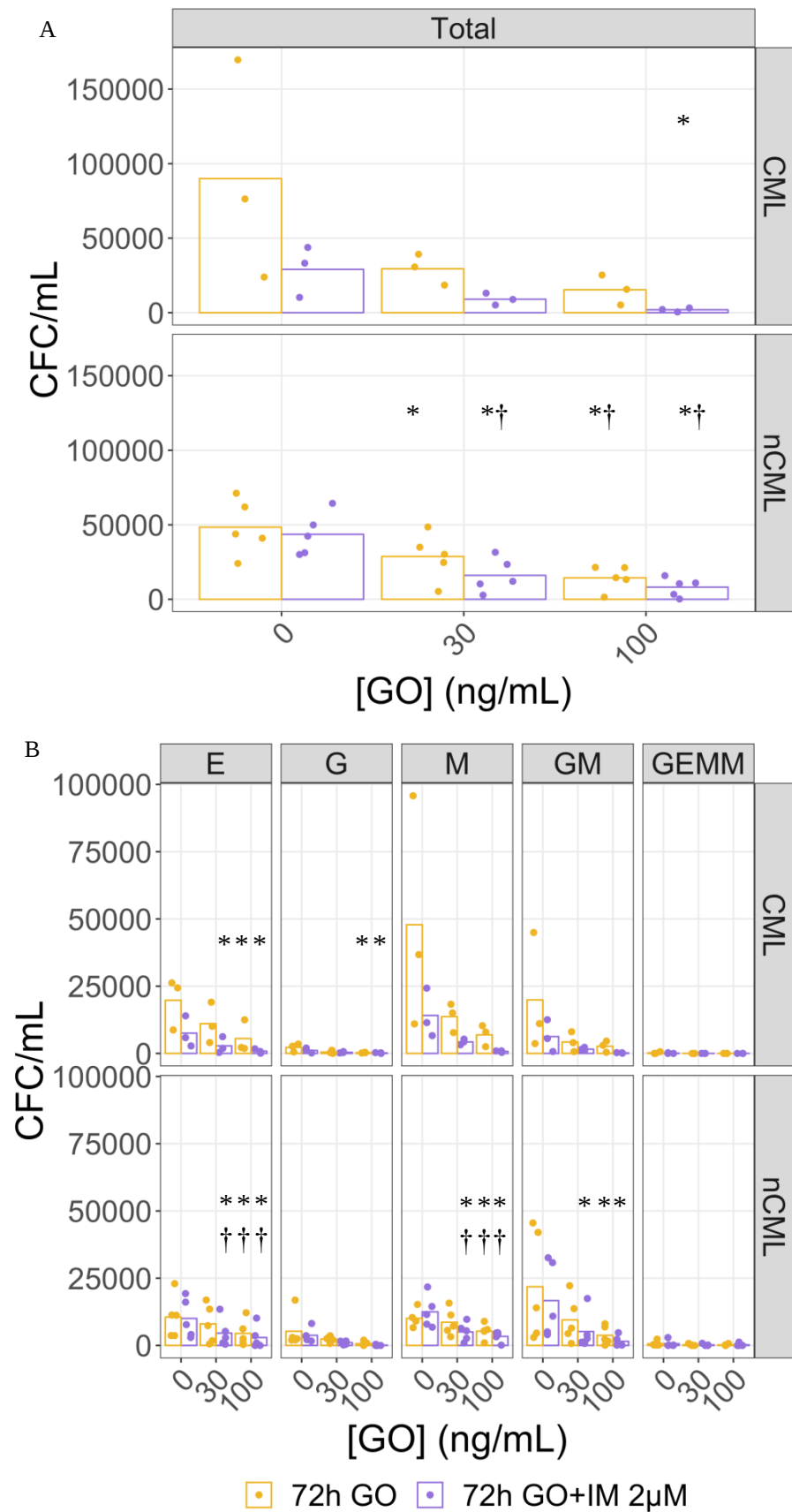


Figure 5-10. CFC concentration after 72h of treatment with GO±IM. (A) Summarised CFC concentrations for each treatment and condition. (B) CFC concentrations for each subtype. Individual dots are the colony counts for each patient and bars are the mean. Significant differences with the NDC are represented with * and significant differences with IM only are represented with †.

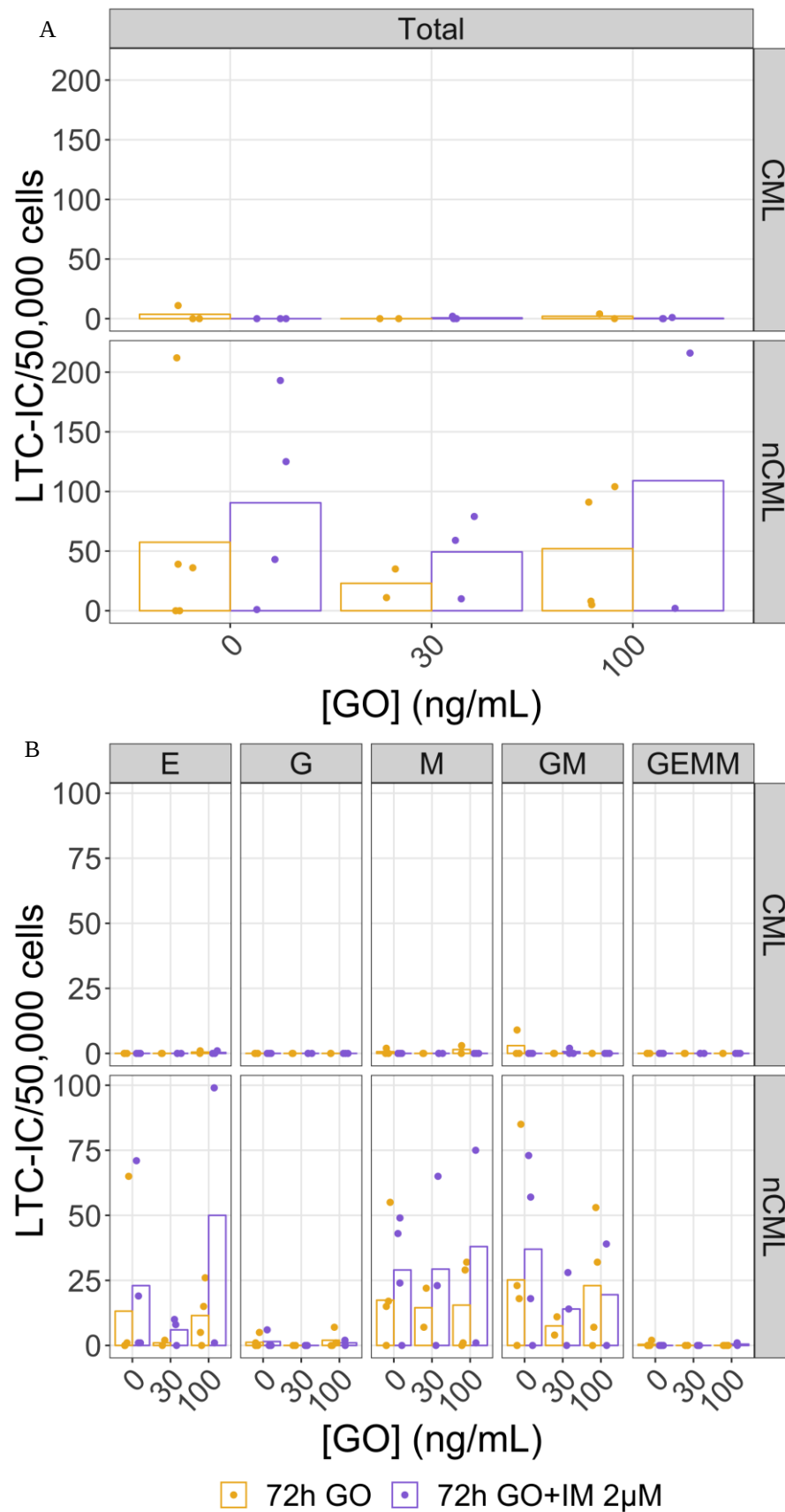


Figure 5-11. Colony counts after 72h of treatment with GO±IM and 6 weeks on culture with stroma cells. (A) Total colony counts for each treatment and condition. (B) Colony counts for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

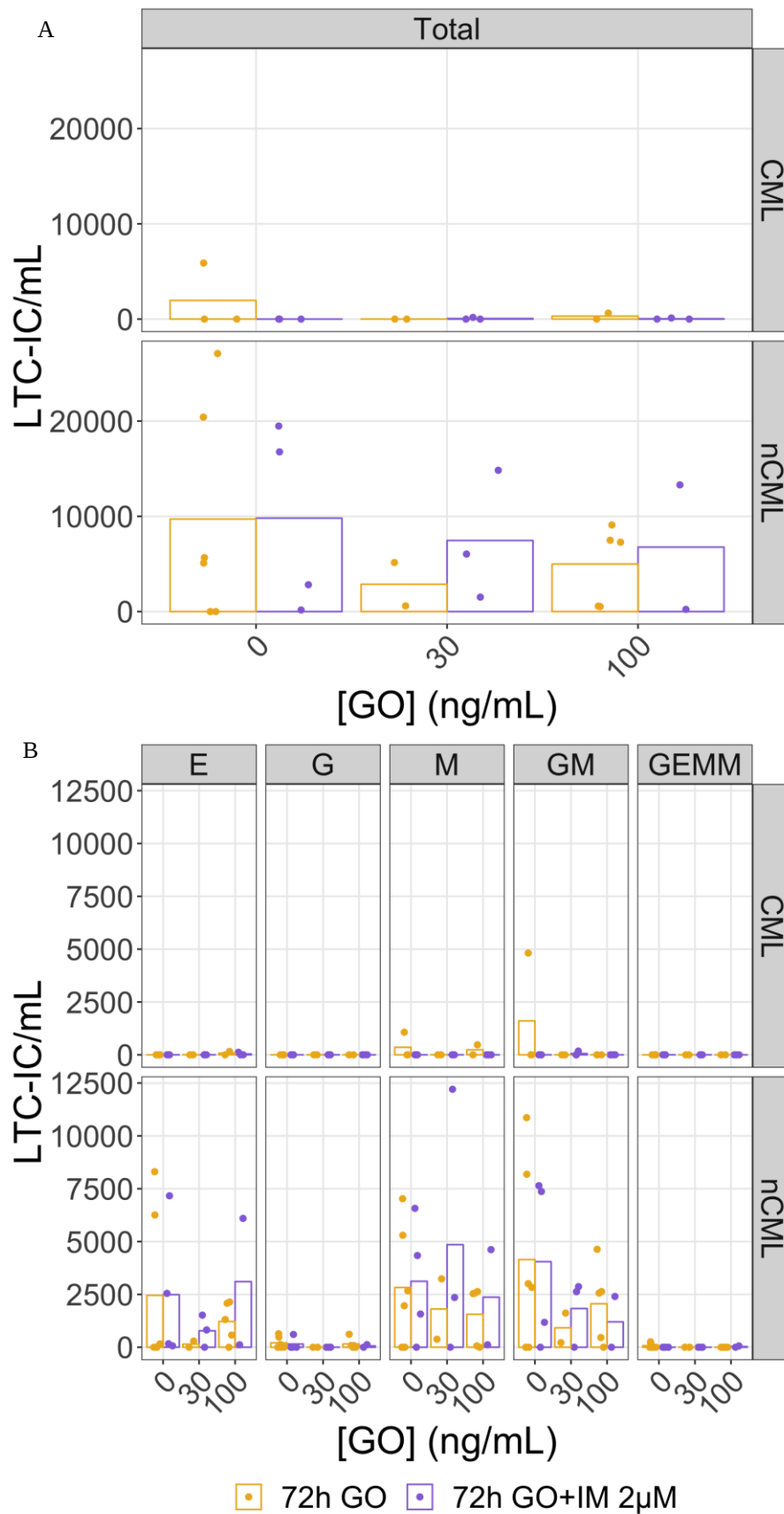


Figure 5-12. CFC concentration after 72h of treatment with GO±IM and 6 weeks on culture with stroma cells. (A) Summarised CFC concentrations for each treatment and condition. (B) CFC concentrations for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

5.2.14 GO increases γ H2AX levels in CML CD34⁺ cells but not in nCML

Similarly to the assay performed in K562 cells, the effect of GO on the formation of double strand breaks in the DNA was assessed by detecting γ H2AX levels by flow cytometry. After subtracting the MFI of the non-stained control from all the samples, the MFI was log-transformed and compared. Paired ANOVA was not able to find statistically significant differences between CML and nCML or between the different treatments (Figure 5-13). However, a modest and constant increase in γ H2AX could be observed in GO alone in CML, which has a significant positive correlation as tested by Spearman test (Spearman's $\rho=0.47$, $p=0.047$), suggesting that the levels of γ H2AX in CML increase with increasing concentrations of GO. Although the combination treatment seemed to also increase in a concentration dependent manner, the effect is more heterogeneous than in the single treatment. The nCML patients seemed not to have increasing levels of γ H2AX with increasing concentrations of GO.

5.2.15 GO promotes cell cycle entry in CML cells while increasing the proportion of quiescent cells in nCML after 72h of treatment

Undivided CML cells are known to persist TKI treatment better than cycling cells (Graham et al., 2002). In order to investigate if GO had any effect on the cell cycle that might affect long-term CML treatment, the percentages of CML cells in the different phases of the cell cycle were analysed by flow cytometry by detecting KI67 and the amount of DNA per cells (using DRAQ7 as DNA stain).

Observation of the data suggests that GO as single agent reduced the proportion of quiescent (G_0) CML cells and increased the proportion of CML cycling cells (S/ G_2 /M) while it increased the proportion of nCML cells in G_0 and decreased the proportion of cells in G_1 (Figure 5-14). As in other experiments, the small power did not allow to detect statistically significant changes between the different concentrations. However, it was possible to detect positive correlation between the percentage of CML cells in S/ G_2 /M and the logarithm of the concentration of GO for the single agent arm (Spearman's $\rho=0.55$, $p=0.018$), while observing a negative correlation between the percentage of nCML cells in G_1 and the logarithm of the concentration of GO for the single agent arm (Spearman's $\rho=-0.38$, $p=0.037$).

Direct comparison of CML and nCML showed a significant difference for the percentage of cells in G_0 when treated with 1000ng/mL of GO ($p=0.011$) and for the percentage of cells in S/ G_2 /M for 30 ($p=0.011$), 100, 300 and 1000ng/mL of GO (all $p<0.01$). The

combination treatment seemed to have a more heterogeneous effect on the cell cycle than the single treatment with IM seeming to antagonise the effect of GO in CML.

The increase in the proportion of quiescent nCML cells might be caused by a biased targeting of GO towards myeloid progenitor cells, which would express higher amounts of CD33 on the cell surface and therefore, an enrichment on more primitive HSCs. This could also explain that this observation has not been repeated in the subsequent sequential treatments, which were performed over 6 days (instead of 3) and a higher proportion of the progenitor cells might have exhausted.

5.2.16 GO reduces the number of CML CD34⁺ cells after 72h of culture followed by 72h of treatment and its effect is additive to a 72h pre-treatment with IM

Currently, most CML patients are on TKI treatment, so the potential introduction of GO into the clinical practice would most probably be for patients that are already on TKI treatment. In order to investigate if TKI treatment improves the response to GO, a sequential treatment of 72h of IM 2 μ M or no-treatment were followed by 72h of GO, the minimum time needed to observe a decrease in cell number in cell lines (section 5.2.1).

It was found that the IC₅₀ of GO after 72h of culture with no drugs and 72h with GO was 213.74 $\pm$ 64.8ng/mL for CML and 700.2 $\pm$ 38.6ng/mL for nCML (p<0.001). When the cells were treated with IM 2 μ M for 72h before the 72h treatment with GO the IC₅₀ was found to be 341.8 $\pm$ 16ng/mL for CML and 1,045.3 $\pm$ 38.6ng/mL for nCML (p<0.001). No statistically significant differences were found between the single treatment and the combination in any of the conditions (CML/nCML). This shows that GO is more effective on CML than in nCML in this treatment regimen as well, with nCML having over 3 times higher IC₅₀s in both the single treatment and the combination (Figure 5-15).

Using Bliss equation (Bliss, 1939) no differences were found between the observed and the expected effects in the combination treatment both in CML and nCML, suggesting that both drugs have additive effects. This was expected for two drugs that act through different mechanisms, such as IM and GO.

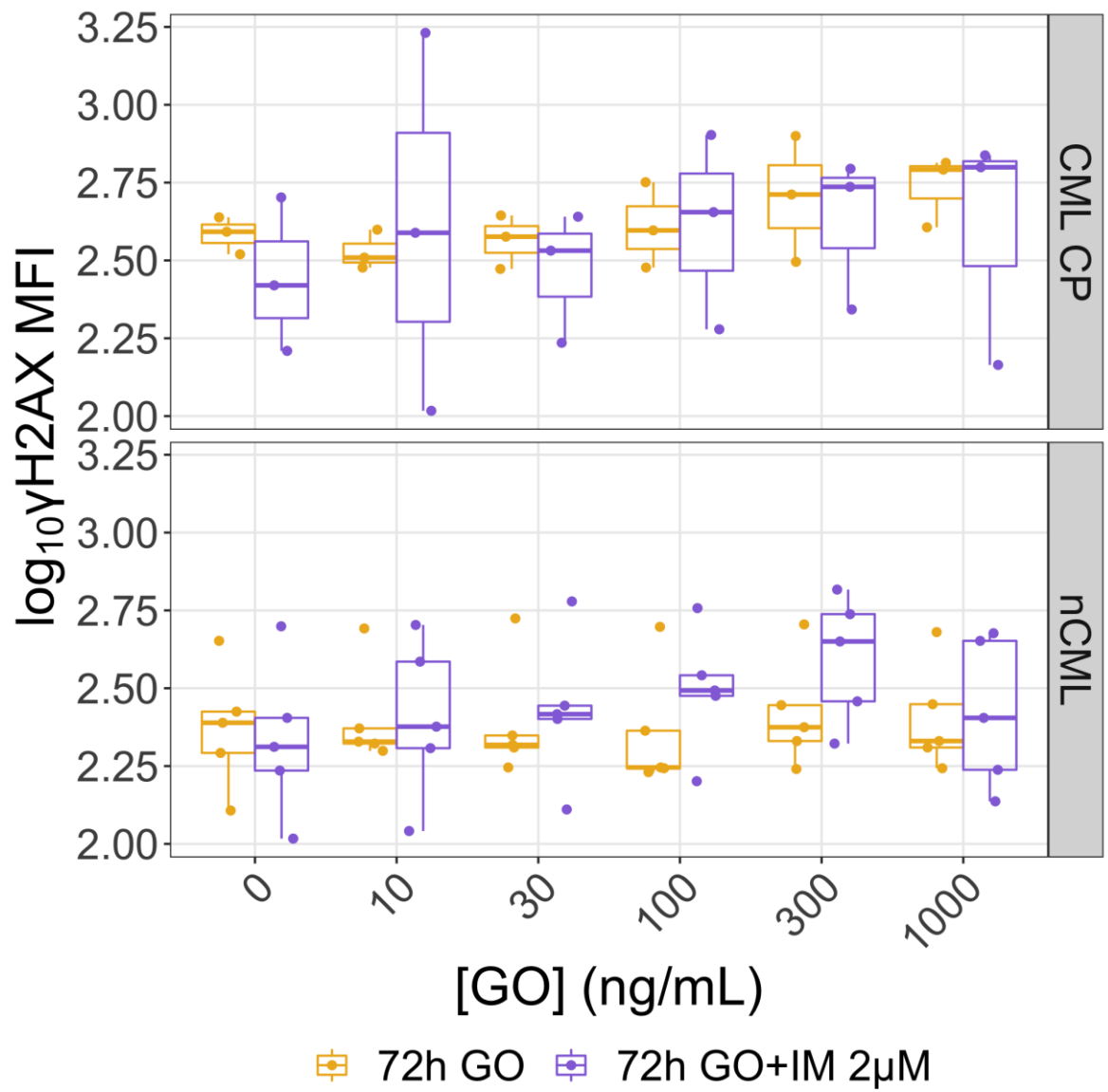


Figure 5-13. Levels of γ H2AX increases in CML after 72h of treatment with GO. CML cells seem to increase the levels of γ H2AX in response to increasing concentrations of GO. Central line in the boxes represents the median of the values. The dots represent the individual samples analysed in the experiment. Replicates: $n_{\text{CML}}=3$, $n_{\text{nCML}}=5$.

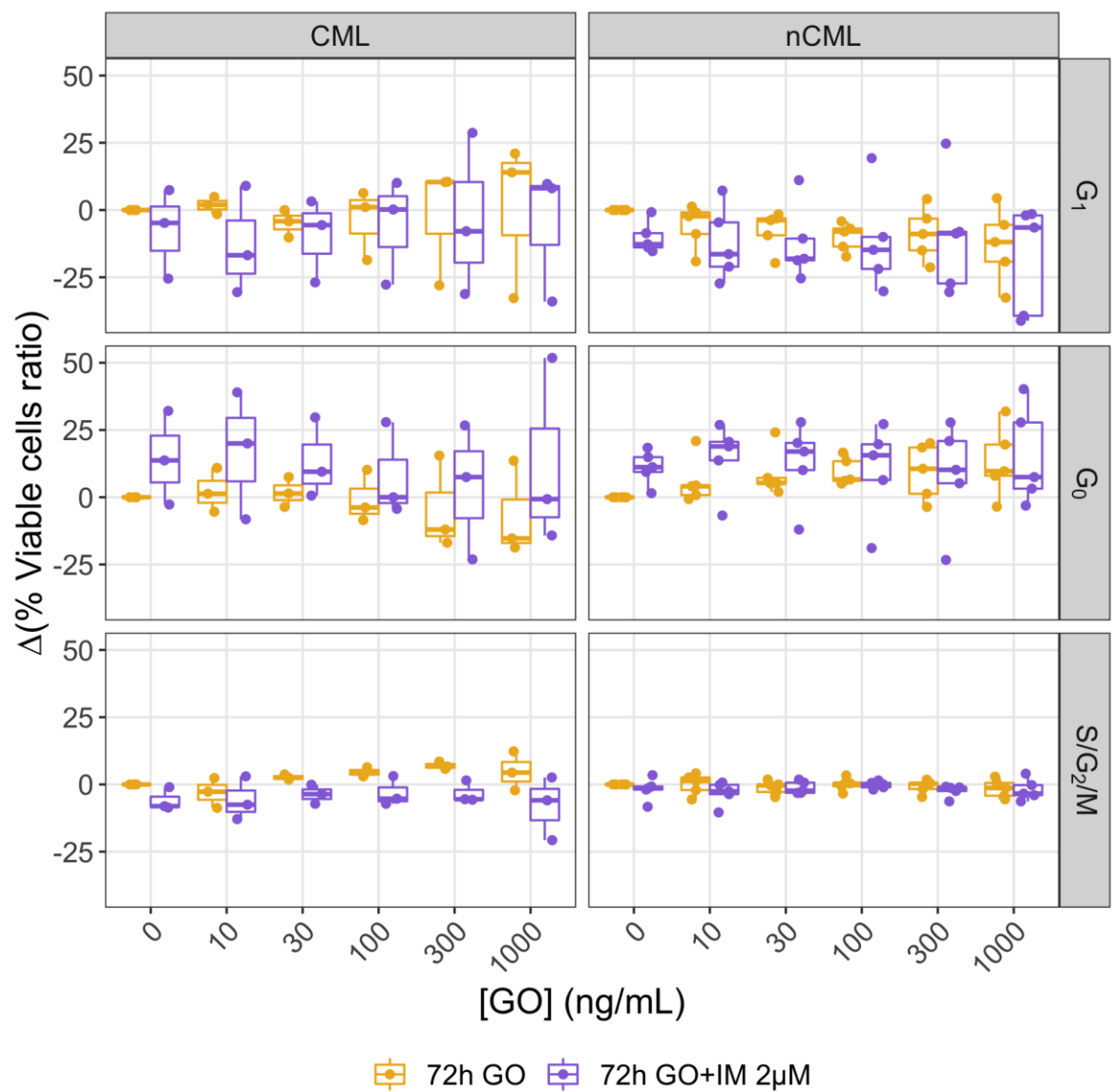


Figure 5-14. Differences on the percentage of cells in each cell cycle phase compared with the NDC. A trend towards a more proliferative phenotype could be observed in CML cells while nCML increased the percentage of quiescent cells. Central line in the boxes represents the median of the values. The dots represent the individual samples analysed in the experiment. Replicates: $n_{\text{CML}}=3$, $n_{\text{nCML}}=5$.

5.2.17 Pre-treatment with IM protects CML cells from entering apoptosis

Pre-treatment with IM did not have any effect in the concentration needed to reduce the percentage of Annexin V⁺ DAPI⁺ cells to half of the NDC in nCML cells ($IC_{50\text{-single treatment}}=200.34\pm64.9\text{ng/mL}$; $IC_{50\text{-combination}}=207.85\pm61.25\text{ng/mL}$; Figure 5-16). However, CML cells seemed to require higher concentrations of GO for entering apoptosis when treated before with IM 2 μM for 72h ($p=0.004$; fig; $IC_{50\text{-single treatment}}=103.97\pm49.59\text{ng/mL}$; $IC_{50\text{-combination}}=269.65\pm68.32\text{ng/mL}$). At the same time, IM alone did not have any effect on apoptosis in any of the cells studied. This could suggest that the cells that persist after TKI treatment are more resistant to GO than the treatment naïve ones. This could be driven by CML cells acquiring a more primitive phenotype after IM treatment, as shown in Figure 5-14 and by previous reports (Charaf et al., 2016), which may reduce the binding points of calicheamicin on the DNA.

5.2.18 IM pre-treatment increases CFC concentration while reducing its total number in CML

The number of CFCs present after the 144h treatment was assessed in the same way as for the 72h treatment (section 5.2.12). The number of total colonies counted for nCML patients was significantly higher than in CML for the NDC ($p=0.006$), 30ng/mL of GO ($p=0.02$) and 100ng/mL of GO ($p=0.04$; Figure 5-17A) but not for the arms treated with IM during the first 72h. This could be explained by a non-significant increase in the colony counts in the CML cells treated with IM compared with the cells that were treated only with GO (Figure 5-17A). A higher number of G and GM colonies was observed in nCML compared with CML for the NDC ($p_G=0.006$; $p_{GM}=0.008$) and IM alone ($p_G=0.042$; $p_{GM}=0.028$; Figure 5-17B).

The total number of CML colonies was significantly higher in the IM alone arm compared with the 100ng/mL of GO without IM ($p=0.011$) and with IM ($p=0.028$). The number of CML E colonies was significantly lower when compared with the only IM treated cells in the 30ng/mL ($p=0.033$) and 100ng/mL GO alone ($p=0.010$) and 100ng/mL of GO after IM treatment groups ($p=0.024$).

The total number of colonies in nCML was significantly lower when compared with the NDC and IM alone for 100ng/mL of GO alone ($p_{NDC}=0.010$; $p_{IM}=0.005$), and 30ng/mL ($p_{NDC}=0.020$; $p_{IM}=0.009$) and 100ng/mL of GO after IM treatment ($p_{NDC}=0.003$; $p_{IM}=0.001$). The number of nCML E colonies was significantly reduced in the 100ng/mL of GO after IM treatment when compared with IM alone ($p=0.025$). The number of nCML

G colonies was lower in 100ng/mL of GO with and without previous IM treatment when compared with NDC ($p_{\text{single100}}=0.034$; $p_{\text{combination100}}=0.039$) and in 100ng/mL of GO without IM treatment, and 30ng/mL and 100ng/mL of GO after IM treatment when compared with IM alone ($p_{\text{single100}}=0.020$; $p_{\text{combination30}}=0.030$; $p_{\text{combination100}}=0.022$). The number of nCML GM colonies was lower in the 100ng/mL of GO after IM treatment arm when compared with NDC ($p=0.040$) and in the 30ng/mL of GO after IM treatment arm when compared with NDC ($p=0.014$) and IM alone ($p=0.033$).

When the number of CFCs was corrected by the concentration of cells in the culture (trypan blue negative cells) the total number of CFCs in the NDC arm was significantly higher in CML compared with nCML ($p=0.002$). It was also possible to observe a non-significant decrease in the number of CFCs when the cells were previously treated with IM in the GO-treated cells (Figure 5-18A).

The number of total and E CML CFCs was higher in the NDC compared with 100ng/mL of GO without ($p_{\text{total}}=0.016$; $p_E=0.023$) and with previous IM treatment ($p_{\text{total}}=0.005$; $p_E=0.013$). All treated arms had reduced number of GM CFCs when compared with the NDC ($p_{\text{IM}}=0.039$; $p_{\text{single-100}}=0.007$; $p_{\text{single-30}}=0.049$; $p_{\text{combination-100}}=0.003$; $p_{\text{combination-30}}=0.024$). All the CFC concentrations can be found in Figure 5-18.

All nCML GO treated arms had significantly lower number of CFCs compared with both the NDC ($p_{\text{single-100}}<0.001$; $p_{\text{single-30}}=0.016$; $p_{\text{combination-100}}<0.001$; $p_{\text{combination-30}}=0.003$) and IM ($p_{\text{single-100}}<0.001$; $p_{\text{single-30}}=0.049$; $p_{\text{combination-100}}<0.001$; $p_{\text{combination-30}}=0.011$) alone, while no difference in the amount of CFCs was observed between these two arms. The number of E nCML CFCs was reduced in the 100ng/mL of GO without previous IM treatment arm when compared with the NDC ($p=0.026$) and in both 100ng/mL of GO arms ($\pm$ IM) when compared with IM alone ($p_{\text{single-100}}=0.017$; $p_{\text{combination-100}}=0.048$). In the nCML patients the number of G CFCs was also significantly reduced in both 100ng/mL of GO arms when compared with the NDC ($p_{\text{single-100}}=0.016$; $p_{\text{combination-100}}=0.030$) and IM alone ($p_{\text{single-100}}=0.028$; $p_{\text{combination-100}}=0.049$) the 30ng/mL of GO after IM treatment arm had a reduced number of G CFCs when compared with the NDC ($p=0.035$). The number of M CFCs was reduced in nCML patients in the two 100ng/mL of GO arms ($\pm$ IM) when compared with the NDC ($p_{\text{single-100}}=0.042$; $p_{\text{combination-100}}=0.013$) and in the 100ng/mL of GO after IM treatment arm when compared with IM alone ($p=0.016$). Similarly to the total number of CFCs, all the nCML GO treated arms had significantly lower number of CFCs than the NDC ($p_{\text{single-100}}=0.003$; $p_{\text{single-30}}=0.022$; $p_{\text{combination-100}}=0.002$; $p_{\text{combination-30}}=0.002$) and all GO-treated arms but the 30ng/mL of GO single treatment had reduced number of CFCs

than IM ($p_{\text{single-100}}=0.031$; $p_{\text{combination-100}}=0.019$; $p_{\text{combination-30}}=0.019$) alone. All the CFC concentrations can be found in Figure 5-18.

As shown in this section, a higher number of nCML comparisons ended up in statistically significant differences than in the case of CML. This happened even with smaller magnitude effects in nCML, potentially because of the bigger heterogeneity of CML compared with nCML and the lower number of CML patients analysed ($n_{\text{nCML}}=7$; $n_{\text{CML}}=4$).

5.2.19 No differences in the number of LTC-ICs were found in GO treated cells previously treated with IM 2 μ M for 72h

As mentioned before, LTC-IC cultures are a challenging assay with many of the samples having very small if any colonies and there is a high inter-patient variability. The number of LTC-ICs detected in the $\pm$ IM 2 μ M for 72h followed by GO 72h was small and neither the total number of LTC-ICs counted (Figure 5-19) nor the estimated original concentration (Figure 5-20) were found to have statistically significant differences between the different treatment arms or conditions (CML/nCML).

5.2.20 Pre-treatment with IM does not affect the level of γ H2Ax, which increases with increasing concentrations of GO

Similarly to the observations of the 72h treatment, no statistically significant differences were found between the different concentrations of GO and/or IM or between CML and nCML. However, it was possible to observe a constant increase in the levels of γ H2AX in nCML for the GO single agent samples and in CML for both GO alone and the combination with IM (Figure 5-21). This was especially clear between 100 and 1000ng/mL of GO in CML. In order to test if there was any correlation, Spearman's rank correlation tested was applied between the levels of γ H2AX detected and the base 10 logarithm of the GO concentration for each treatment arm and condition. CML samples with no previous IM treatment were slightly correlated with the concentration of GO (Spearman's $\rho=0.402$, $p=0.052$) while the samples pre-treated with IM had a higher correlation with the concentration of GO (Spearman's $\rho=0.642$, $p<0.001$). Correlation became stronger for the interval [100, 1000], with CML not pre-treated with IM getting a Spearman's ρ of 0.650 ($p=0.022$) and the samples pre-treated with IM a Spearman's ρ of 0.857 ($p<0.001$). The nCML samples were also slightly correlated with the GO concentration when they were not pre-treated with IM (Spearman's $\rho=0.448$, $p=0.028$) but not when they were pre-treated with IM or in the interval [100, 1000].

5.2.21 A delayed start on GO treatment induces increasing percentage of nCML cells entering S/G₂/M phases of the cell cycle

The small sample size and the heterogeneity of the data, especially CML patients, made it difficult to detect statistically significant changes between the different GO and IM concentrations. The only significant difference detected between GO concentrations was between the difference in the percentage of cells in phases S/G₂/M between 10ng/mL of GO after IM treatment and 1,000ng/mL of GO with no previous IM treatment in nCML (Figure 5-22).

Similarly to what was observed in the 72h treatment regimen, it was possible to observe some consistent trends, especially a decrease of CML cells in G₀ and an increase of CML cells in G₁. Additionally, a small increase in the percentage of cycling cells (S/G₂/M) could be observed in nCML. The heterogeneity of the CML samples did not allow for detection of a significant correlation but the increase in the percentage of cycling nCML cells was significantly correlated with the logarithm of GO concentration when the cells were not pre-treated with IM (Spearman's $\rho=0.484$, $p=0.017$).

Comparison of CML and nCML revealed differences between the percentages of cells in G₁ phase for the NDC ($p=0.023$), and 10ng/mL ($p=0.049$) and 30ng/mL of GO alone ($p=0.035$). This was due to nCML samples having in overall a higher percentage of cells in G₁ phase than CML but not because of noticeable differences in the percentage of nCML cells in G₁ between different concentrations of GO.

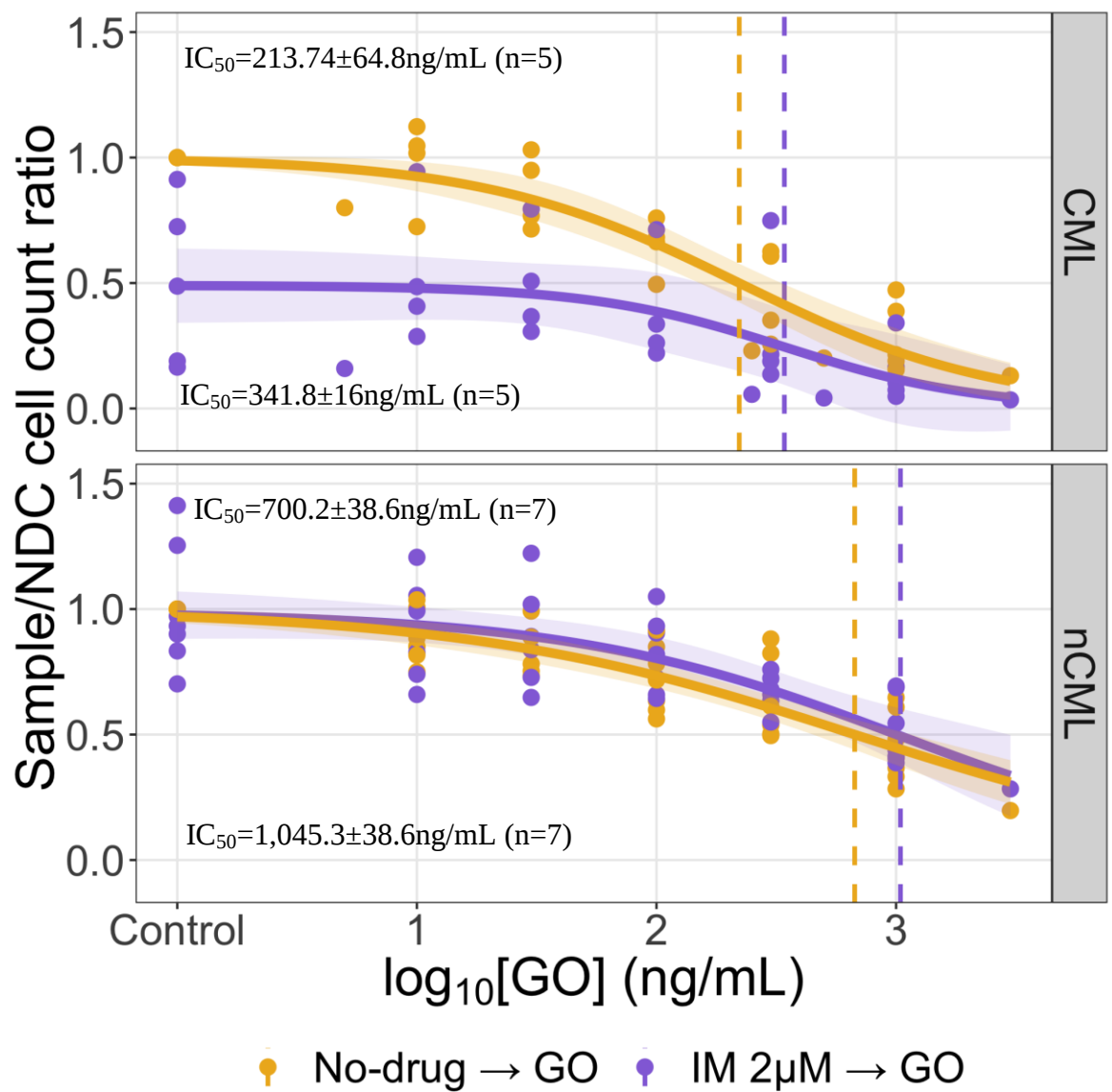


Figure 5-15. Treatment with GO after IM reduces CML cell number. GO treatment was able to reduce the number of CML cells at lower concentrations than nCML, as can be seen in the two different IC_{50} s. Also, GO further reduced the number of CML cells when combined with IM but had a limited effect on nCML cells at concentrations lower than 300 ng/mL. The IC_{50} s at the top of each panel correspond to the GO single treatment while the IC_{50} s at the bottom correspond to the combination treatment. Response to the treatment is represented with the solid line with 95% confidence interval (shadowed area). The dashed lines represent the IC_{50} s. Individual dots represent individual observations.

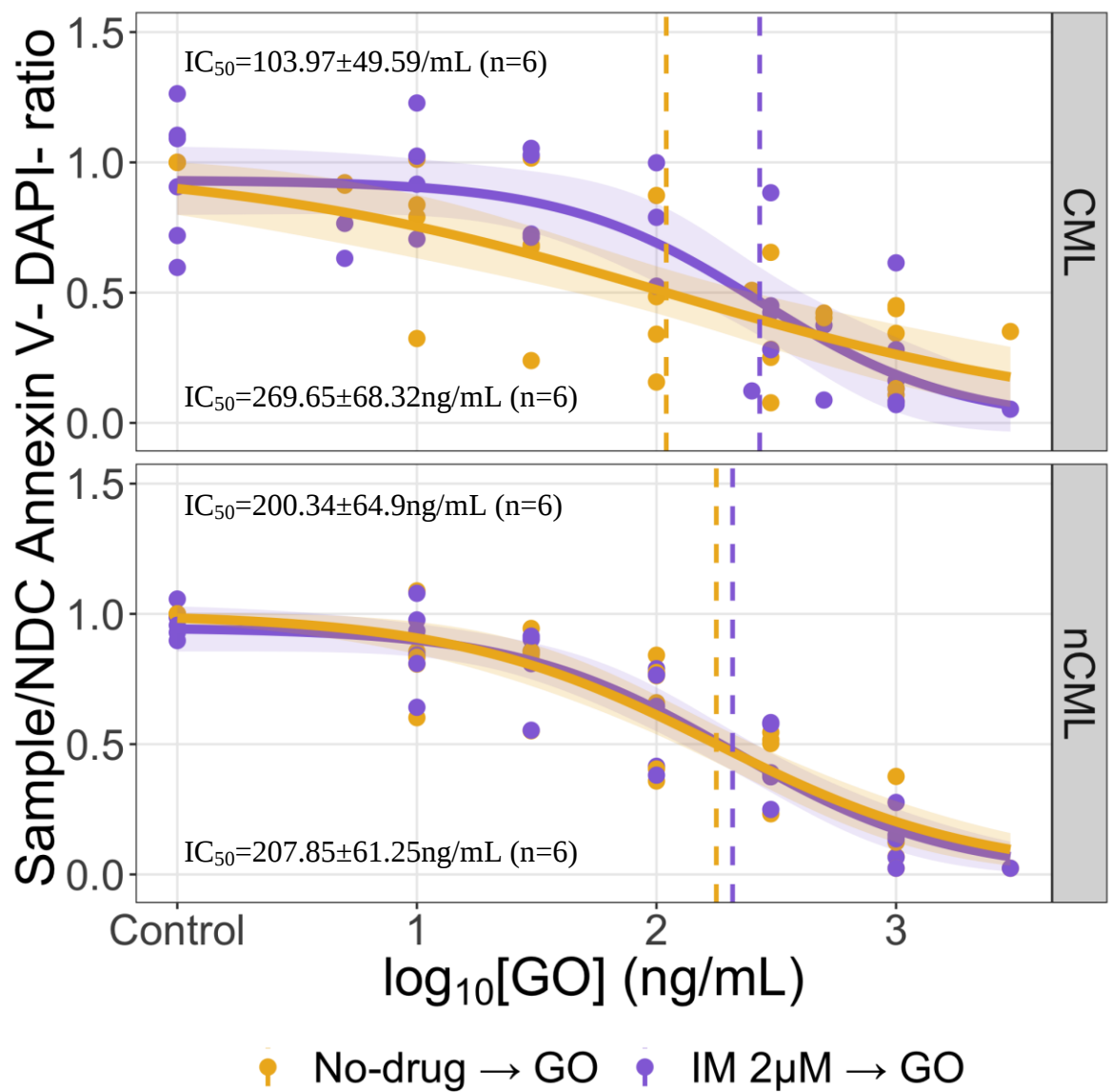


Figure 5-16. GO after IM/no treatment reduced the percentage of viable cells in both CML and nCML. The percentage of viable cells (Annexin V⁺ DAPI⁺) was reduced similarly in CML and nCML as it was between GO single treatment and its combination with IM. The IC₅₀s at the top of each panel correspond to the GO single treatment while the IC₅₀s at the bottom correspond to the combination treatment. Response to the treatment is represented with the solid line with 95% confidence interval (shadowed area). The dashed lines represent the IC₅₀s. Individual dots represent individual observations.

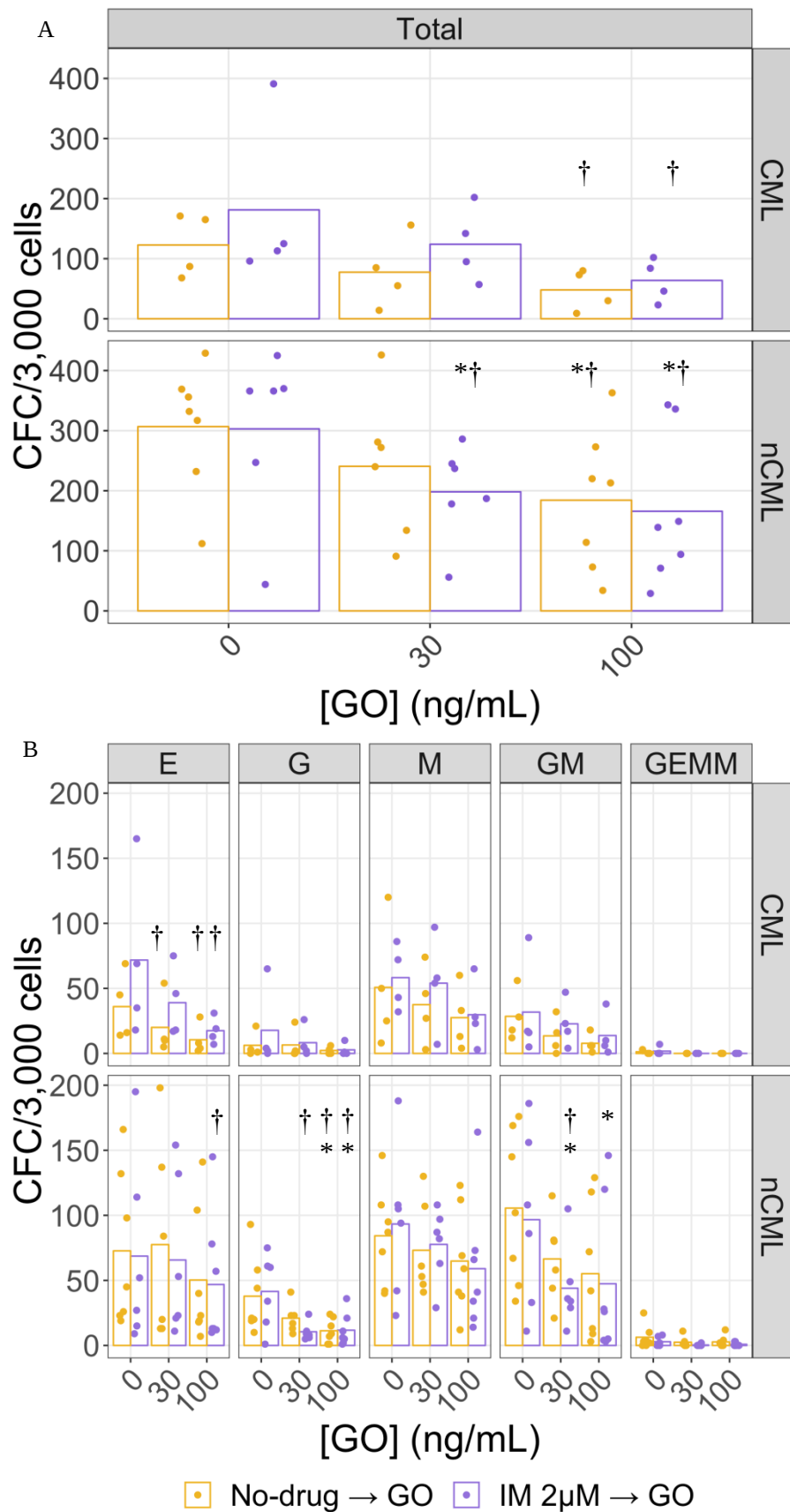


Figure 5-17. Colony counts after sequential treatment IM/no drug → GO. (A) Total colony counts for each treatment and condition. IM enrich CML samples in CFCs. (B) Colony counts for each subtype. Individual dots are the colony counts for each patient and bars are the mean. Significant differences with the NDC are represented with * and significant differences with IM only are represented with †.

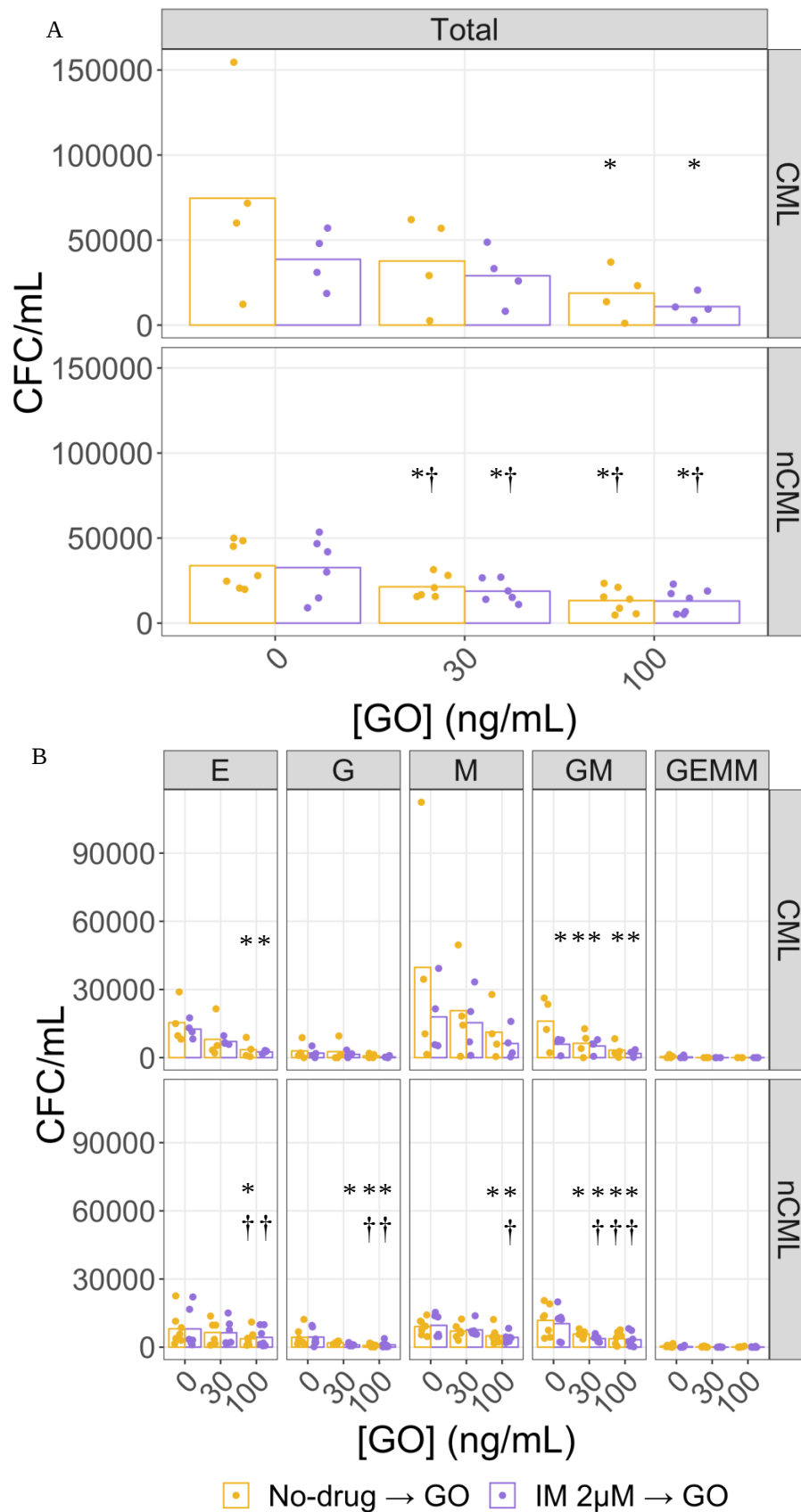


Figure 5-18. CFC concentration after sequential treatment IM/no drug → GO. (A) Summarised CFC concentrations for each treatment and condition. (B) CFC concentrations for each subtype. Individual dots are the colony counts for each patient and bars are the mean. Significant differences with the NDC are represented with * and significant differences with IM only are represented with †.

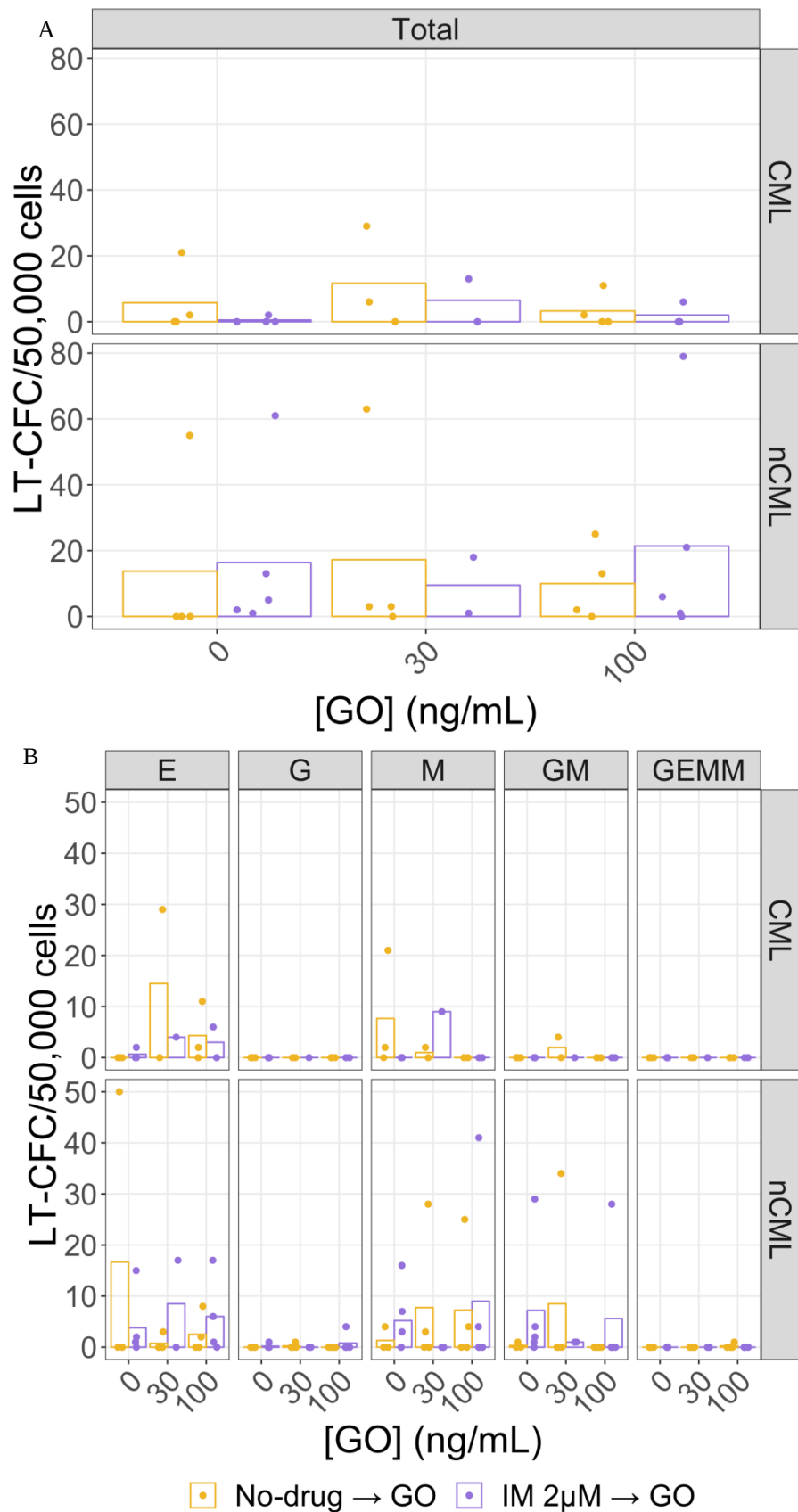


Figure 5-19. Colony counts after sequential treatment IM/no drug → GO and 6 weeks on culture with stroma cells. (A) Total colony counts for each treatment and condition. (B) Colony counts for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

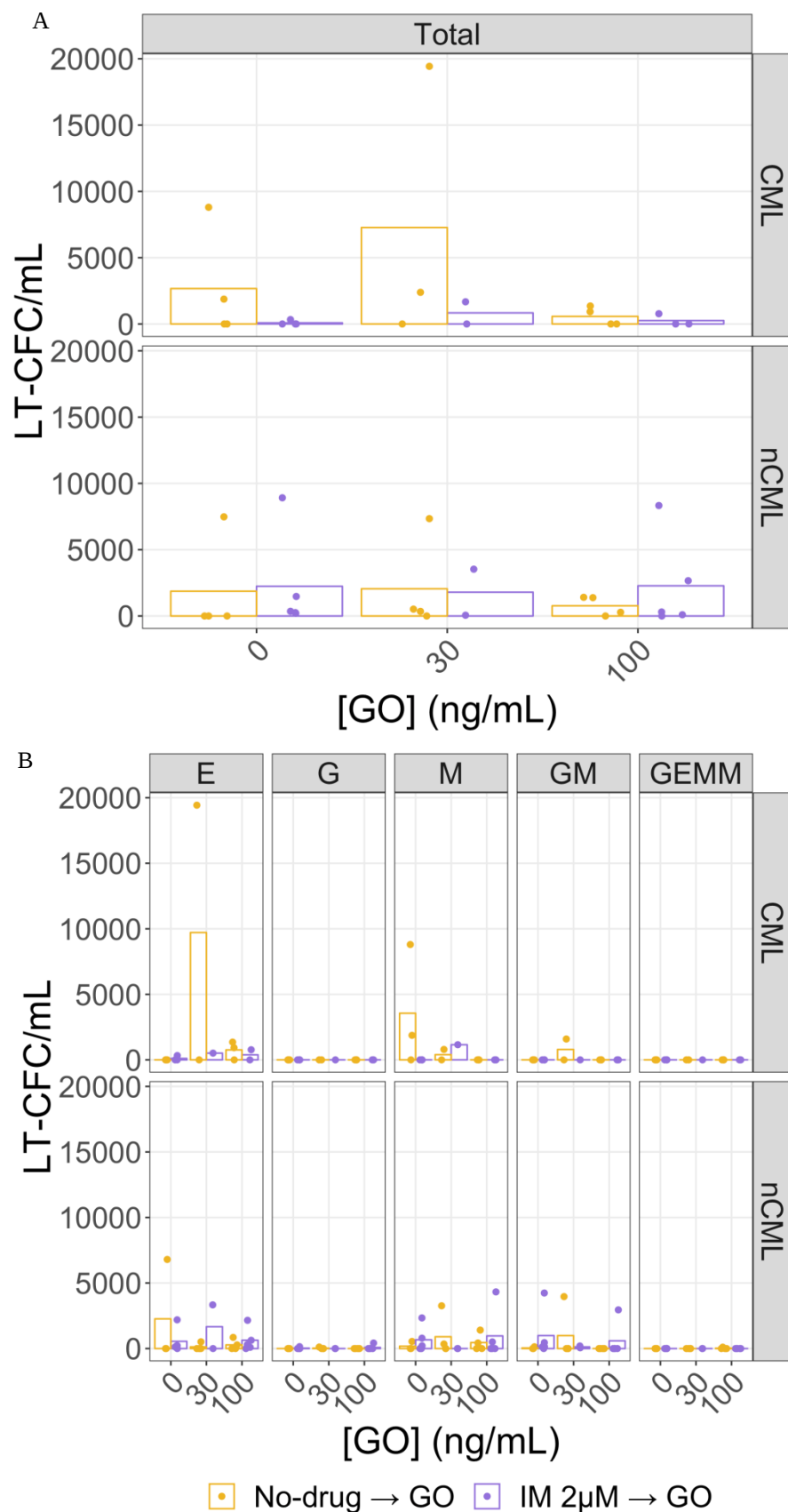


Figure 5-20. CFC concentration after sequential treatment IM/no drug → GO and 6 weeks on culture with stroma cells. (A) Summarised CFC concentrations for each treatment and condition. (B) CFC concentrations for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

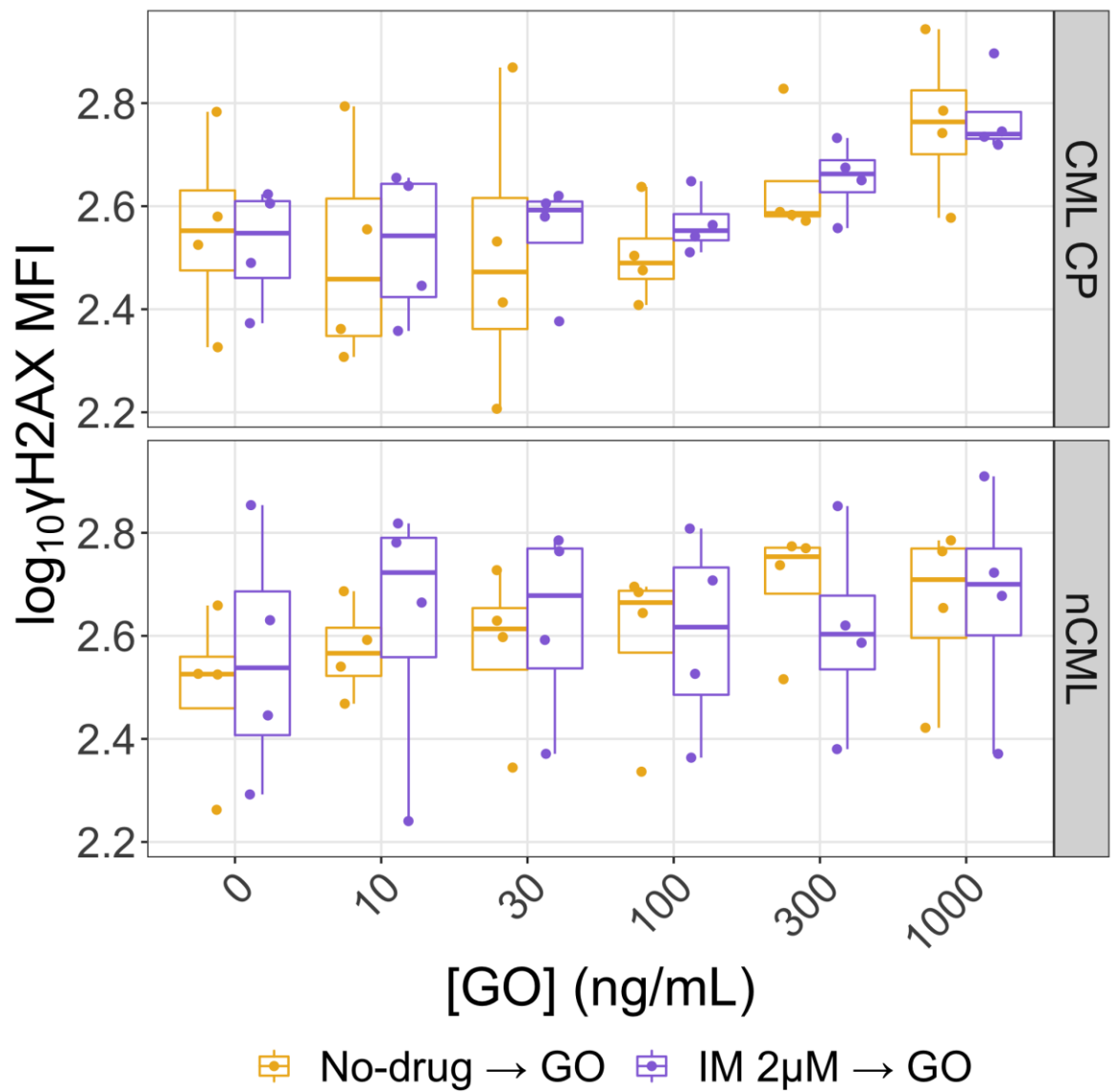


Figure 5-21. Levels of γ H2AX increases in CML and nCML after sequential treatment IM/no drug $\rightarrow$ GO. CML cells seem to increase the levels of γ H2AX in response to increasing concentrations of GO, especially with concentrations between 100 and 1000ng/mL, both when they have been treated previously with IM or not. The nCML samples increase γ H2AX with increasing GO concentrations. The nCML samples previously treated with IM had higher variability and no correlation was found. Central line in the boxes represents the median of the values. The dots represent the individual samples analysed in the experiment. Replicates: $n_{\text{CML}}=4$, $n_{\text{nCML}}=4$.

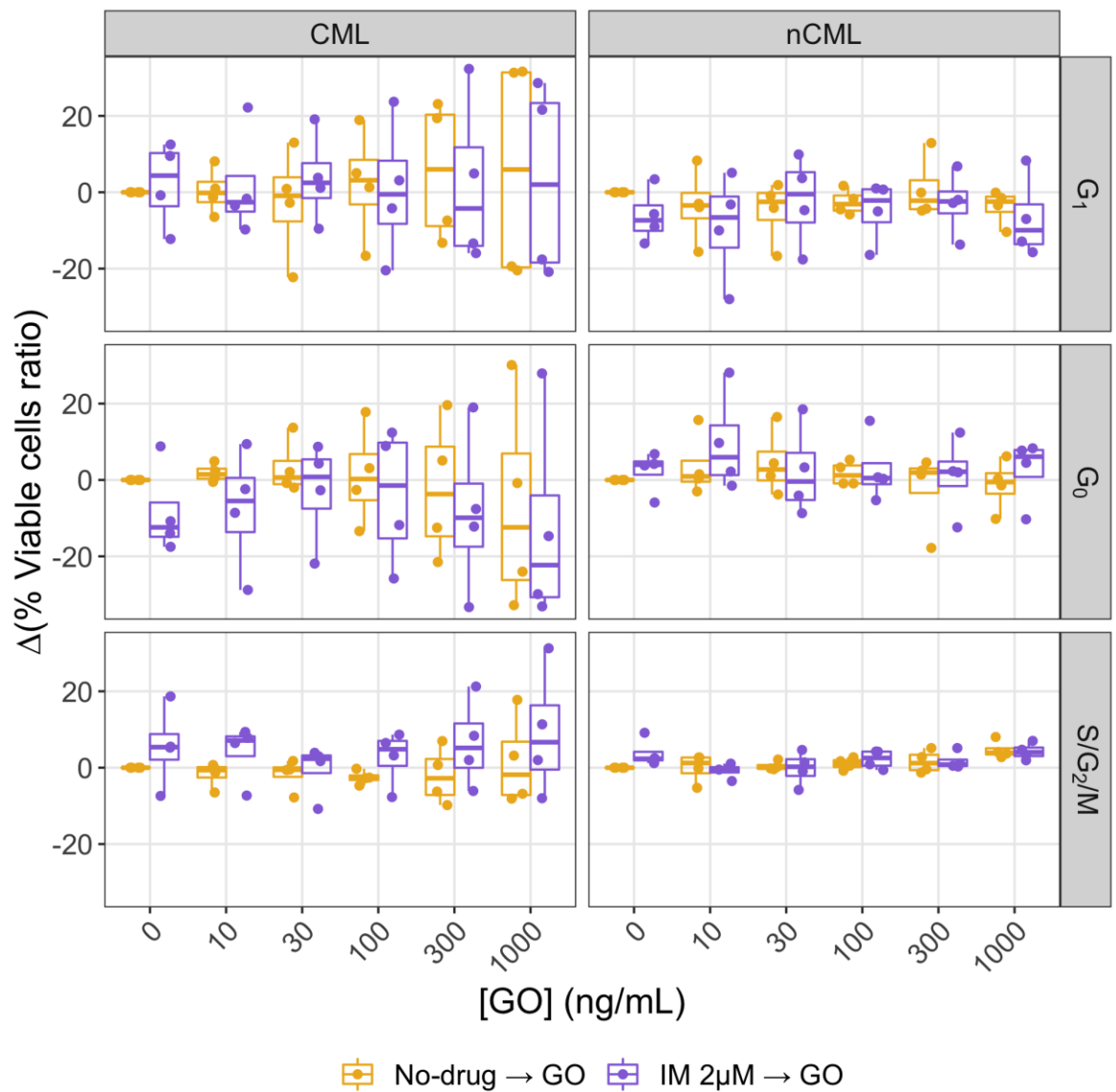


Figure 5-22. Differences on the percentage of cells in each cell cycle phase compared with the NDC after sequential treatment IM/no drug → GO. A trend towards a more proliferative phenotype can be observed in CML and nCML cells. The difference in cells in S/G₂/M phase positively correlated with the logarithm of GO concentration in nCML while in CML a non-significant trend can be observed for a reduction in the percentage of cells in G₀. Central line in the boxes represents the median of the values. The dots represent the individual samples analysed in the experiment. Replicates: $n_{\text{CML}}=4$, $n_{\text{nCML}}=4$.

5.2.22 Treatment with IM after treating CD34⁺ cells with GO synergises with GO in reducing cell number

As a sequential treatment of IM followed by GO was put in place it was thought that testing the reverse regimen could help to understand the underlying interactions between both drugs. Therefore, the cells were treated for 72h with different concentrations of GO and then, allowed to recover for 72h or treated for 72h with IM. The limitation in the number of cells only allowed for analysing 3 CML and 1 nCML patient under this treatment regimen.

The IC₅₀ for GO alone in CML was 74.13±16.36ng/mL but it decreased to 31.91±9.49ng/mL when the cells were treated with IM instead of allowing them to recover (p=0.133). The nCML patient presented an IC₅₀ of 154.75±109.01ng/mL for the single treatment and 69.50±42.81ng/mL for the combination treatment (Figure 5-23).

Bliss equation revealed an average increase of 6.87% in the response from the observed response of the combination of GO and IM to the expected response in CML (p=0.038). In nCML there was also an increase of 6.18% between the observed and expected responses. These values suggest that IM and GO act in a synergistic manner when the cell are first treated with GO and then with IM.

5.2.23 Medium-term effect of GO in percentage of viability can be observed in concentrations equal of higher than 100ng/mL

Observing the percentage of Annexin V⁺ DAPI⁺ cells in the concentration-response curve (Figure 5-24), it is easy to observe that CML cells treated with less than 100ng/mL of GO and cultured for another 72h without any drug treatment fully recovered and presented similar percentages of viable cells than the NDC. On the other hand, cells that were treated with at least 100ng/mL of GO presented higher levels of non-viable cells even after 72h and a reduction in half the percentage of viable cells at 171.60±31.61ng/mL. Treatment with IM prevented the cells from recovering and increased percentages of non-viable cells could be observed at concentrations lower than 100ng/mL, reaching a decrease in half the percentage of viable cells at 54.99±23.02ng/mL. However, this difference was not statistically significant. The nCML cells presented a modest decrease in viability at concentrations less than 100ng/mL and a concentration of 341.27±263.17ng/mL was required for reducing the number of viable cells by half of the NDC. IM did not have any major effect on nCML (IC₅₀=196.76±163.92; p=0.71).

5.2.24 IM treatment after GO treatment had a moderate effect at reducing the total number of CFCs in CML

CFC assays revealed no changes in CFCs counts in the different treatment arms in CML ($p=0.343$) but a small decrease in the CFC count in nCML: from 373 in the NDC to 266 in the 100ng/mL of GO single arm (Figure 5-25A). No statistically significant differences were found in the individual colony subtypes (Figure 5-25B).

When the colony counts were corrected by the cell concentration in the original culture it was possible to observe a non-statistically significant decrease in the number of total CML CFCs with increasing concentrations of GO and, especially a major effect in the 100ng/mL of GO followed by IM treatment (Figure 5-26A). It is also possible to observe a total CFC decrease in nCML. Observing the individual colony subtypes significant decreases in CML G CFCs could be found in all treatment arms when compared with the NDC ($p_{IM}=0.044$; $p_{single-100}=0.010$; $p_{single-30}=0.005$; $p_{combination-100}=0.002$; $p_{combination-30}=0.002$). No other individual subtypes presented statistically significant differences, although the decrease in the number of CFCs was a constant trend in the data (Figure 5-26B).

5.2.25 IM treatment after GO treatment further reduces the number of LTC-ICs in CML

Only the samples from one patient, CML460, developed colonies consistently after 6 weeks of culture. The other samples did not develop colonies (Figure 5-27). CML460 developed mainly E colonies (Figure 5-27A) and it was possible to observe that while IM increased the number of LTC-ICs per cell in the absence of GO it reduced the number of LTC-ICs when used after GO (Figure 5-27). Analysing the data corrected by cell concentration in the original culture it was possible to observe that cells previously treated with GO contain less LTC-ICs when treated with IM afterwards (Figure 5-28).

5.2.26 Treating CML cells with IM that were previously treated with GO increases the levels of γ H2AX

DNA damage response was also assessed for GO followed by IM treatment regimen by the levels of γ H2AX. PGT170810, the nCML patient sample that was treated under this regimen, showed a 50% increase in γ H2AX levels after IM treatment but only very modest increases after low concentrations of GO and reached an increase of 30% with concentrations of 300 and 1000ng/mL of GO (Figure 5-29). Although the non-IM arm in nCML was positively correlated with the logarithm of the concentration of GO, the sample size ($n=1$) did not provide enough power for the correlation test to be significant

(Spearman's $\rho=0.829$, $p=0.058$). The correlation did not improve selecting only the [100, 1000] interval for the concentrations of GO.

The level of γ H2AX seemed to increase in CML at concentrations higher than 100ng/mL when the cells were treated only with GO. IM treatment seemed to reduce the levels of γ H2AX when the cells were not treated previously with GO but when used after GO it reduced the concentrations of GO required to increase the levels of γ H2AX (Figure 5-29). Although these differences were not statistically significant, the constant increase of γ H2AX with increasing concentrations of GO allowed to calculate the correlation values using Spearman's correlation test. The combination treatment had a statistically significant correlation (Spearman's $\rho=0.678$, $p=0.015$) while GO single agent was close to be significant (Spearman's $\rho=0.539$, $p=0.071$). These correlation values improved when only the intervals where the highest observed increases were found: the GO single treatment had an improved correlation when only the interval [100, 1000] was taken in account (Spearman's $\rho=0.837$, $p=0.038$) and the combination treatment when the interval [0, 30] was considered (Spearman's $\rho=0.956$, $p=0.003$).

5.2.27 Treatment with GO reduced the percentage of cells in G_0 72h after stopping the treatment in both CML and nCML and seemed to be independent of IM

In this treatment regimen both CML and nCML followed similar changes in cell cycle after treatment with GO $\pm$ IM. The changes consisted of an increase in the percentage of cells in S/G₂/M and G₁ phases and a decrease in the percentage of cells in G₀ (Figure 5-30). The percentage of CML cells in G₀ was significantly lower than the NDC when the cells were treated with concentrations of GO of 300 or 1000ng/mL ($p<0.05$ for all of them). No statistically significant differences were found in the other phases of cell cycle after correction for multiple testing.

When the differences in the percentages of cells in each phase of the cell cycle were tested for correlation, the differences in the CML cells treated only with GO were correlated with the concentration of GO for G₁ (Spearman's $\rho=0.793$, $p=0.002$), G₀ (Spearman's $\rho=-0.906$, $p<0.001$) and S/G₂/M (Spearman's $\rho=0.821$, $p=0.001$). The combination treatment in CML only had a significant correlation in G₀ (Spearman's $\rho=-0.594$, $p<0.042$). The nCML sample was significantly correlated at both G₀ (Spearman's $\rho=-1$, $p<0.003$) and S/G₂/M phases (Spearman's $\rho=0.943$, $p=0.017$) with the GO single treatment.

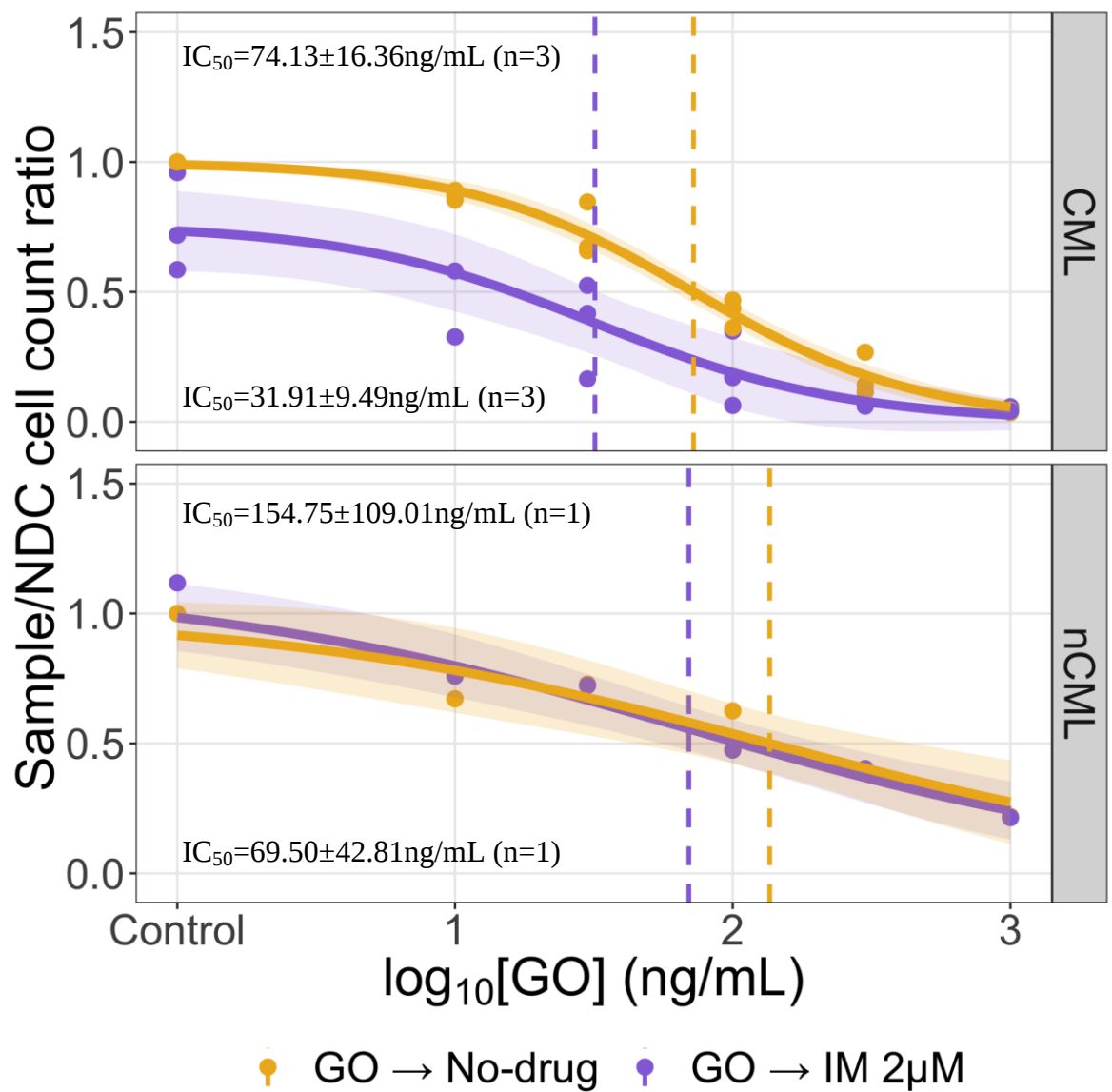


Figure 5-23. Treatment with GO before IM reduces cell number in CML and nCML. GO treatment was able to reduce the number of CML cells at lower concentrations than nCML, as can be seen in the two different IC₅₀s. Also, GO further reduced the number of CML cells when combined with IM while having a milder effect on nCML cells. The IC₅₀s at the top of each panel correspond to the GO single treatment while the IC₅₀s at the bottom correspond to the combination treatment. Response to the treatment is represented with the solid line with 95% confidence interval (shadowed area). The dashed lines represent the IC₅₀s. Individual dots represent individual observations.

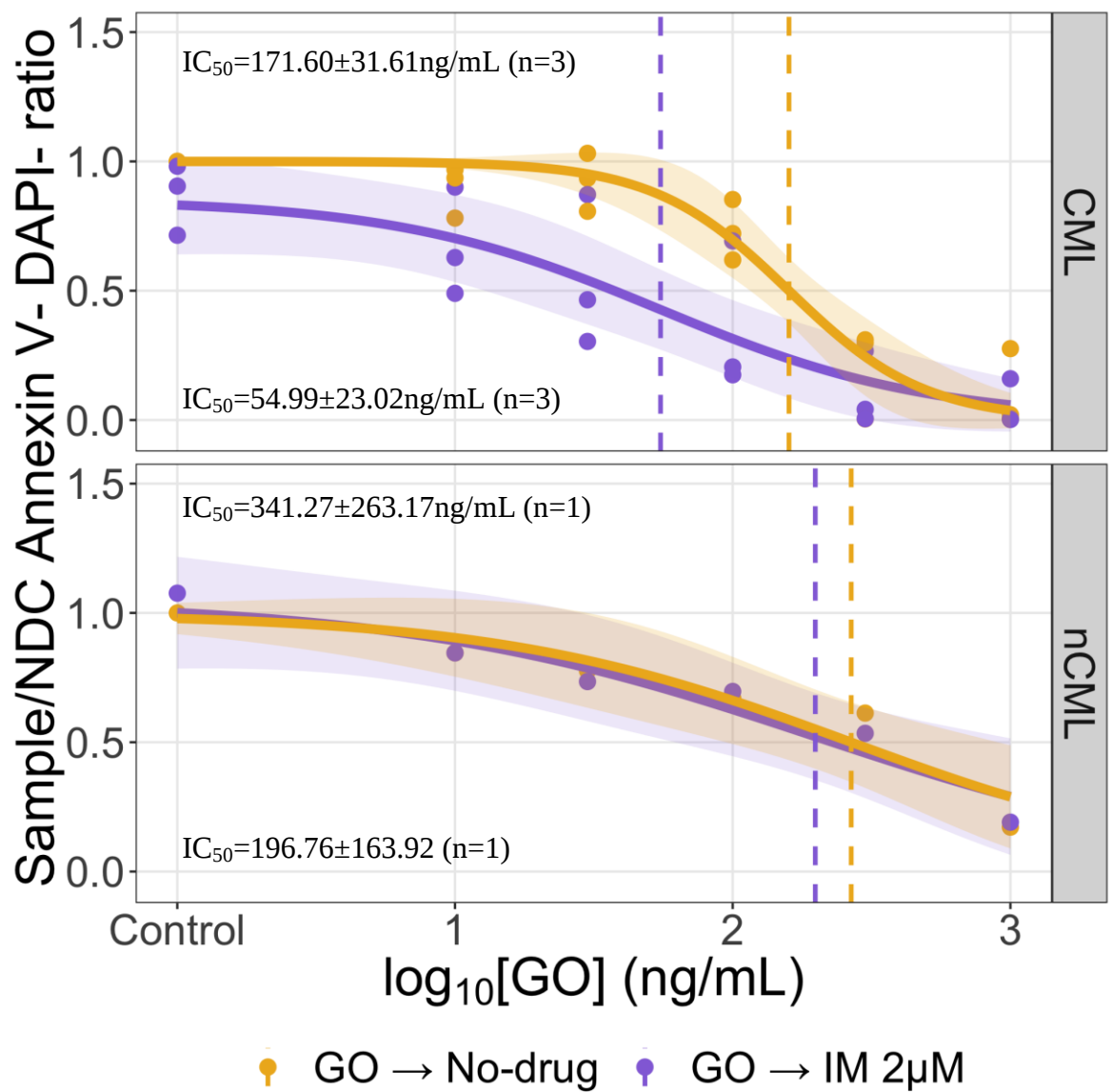


Figure 5-24. Treatment with IM prevented CML cells from recovering after GO treatment. The percentage of viable cells (Annexin V⁻ DAPI⁺) was similar in CML for concentrations of GO lower than 100ng/mL when the cells were not treated with IM after GO but that percentage was reduced when the cells were treated with IM. IM treatment had a minor effect on nCML cells. The IC_{50} s at the top of each panel correspond to the GO single treatment while the IC_{50} s at the bottom correspond to the combination treatment. Response to the treatment is represented with the solid line with 95% confidence interval (shadowed area). The dashed lines represent the IC_{50} s. Individual dots represent individual observations.

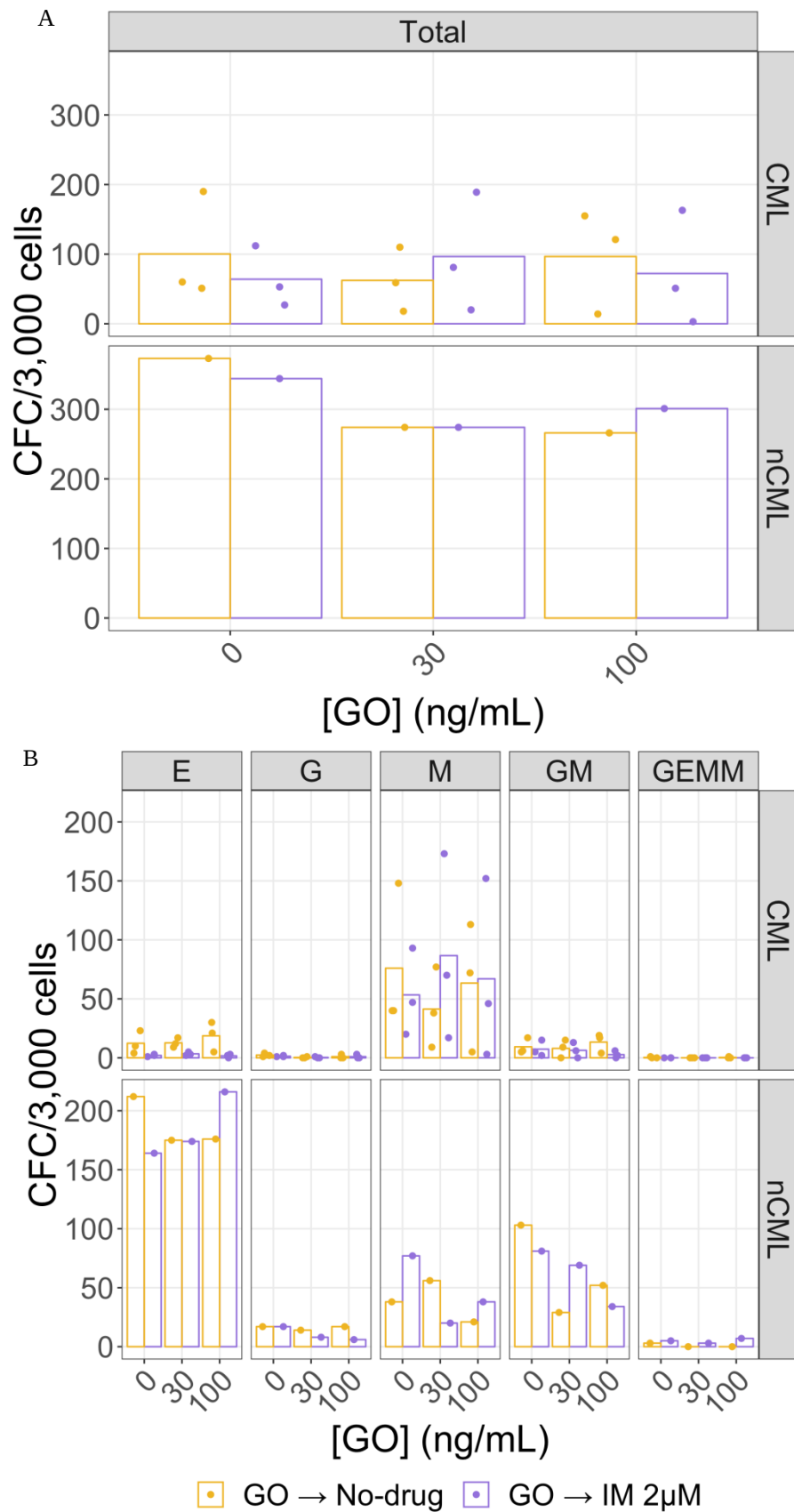


Figure 5-25. Colony counts after sequential treatment GO → IM/no drug. (A) Total colony counts for each treatment and condition. (B) Colony counts for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

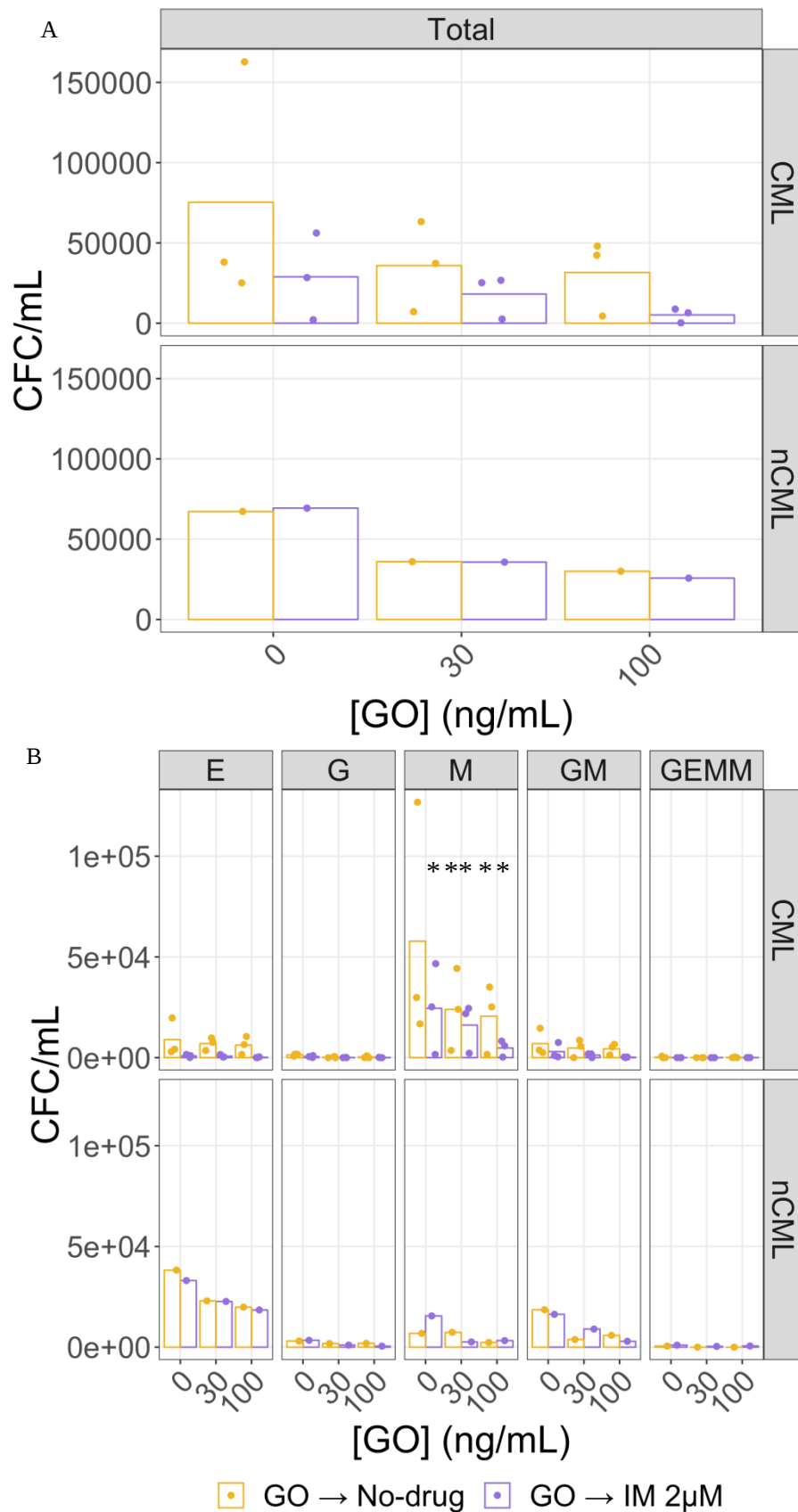


Figure 5-26. CFC concentration after sequential treatment GO → IM/no drug. (A) Summarised CFC concentrations for each treatment and condition. (B) CFC concentrations for each subtype. Individual dots are the colony counts for each patient and bars are the mean. Significant differences with the NDC are represented with *.

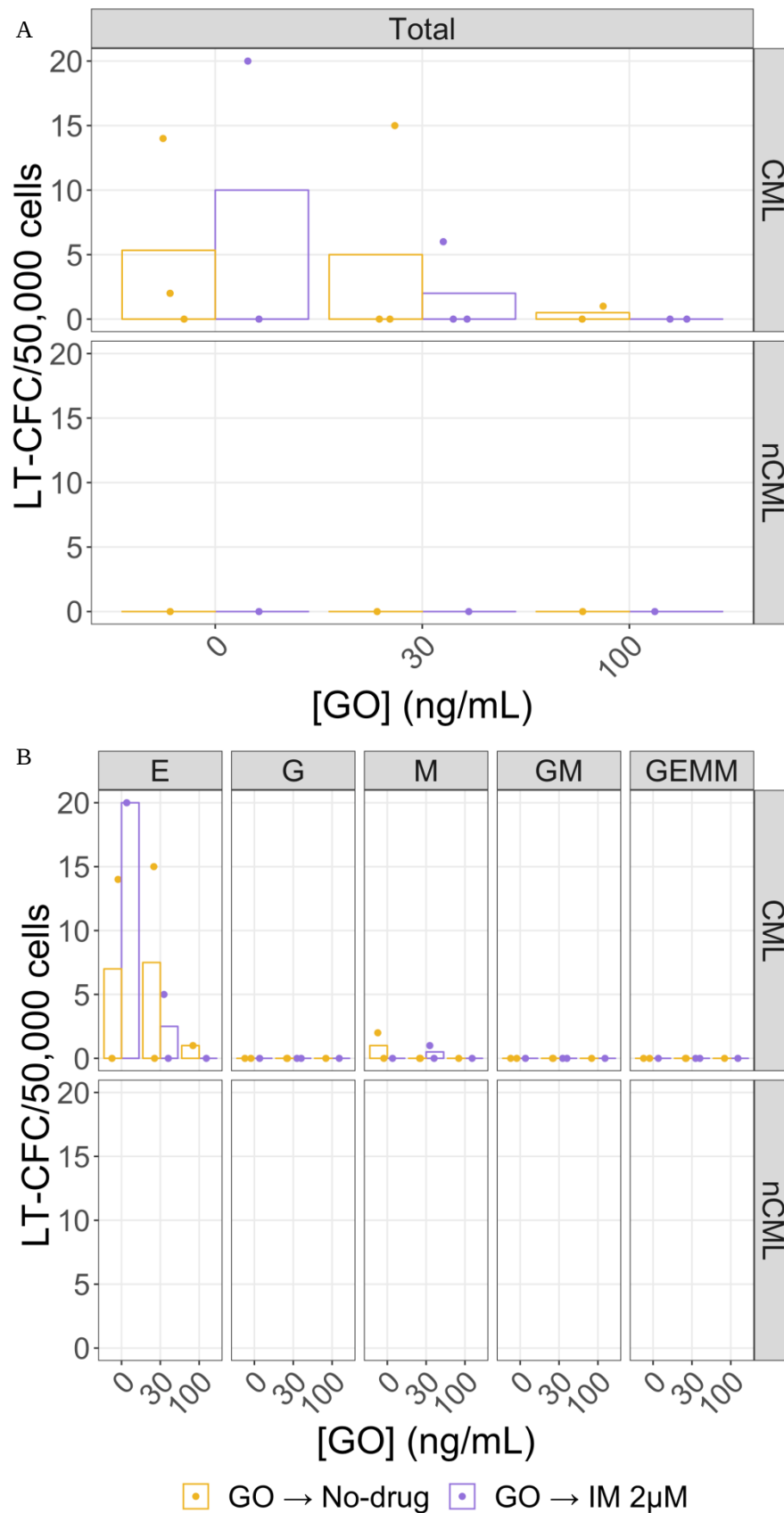


Figure 5-27. Colony counts after sequential treatment GO → IM/no drug and 6 weeks on culture with stroma cells. (A) Total colony counts for each treatment and condition. (B) Colony counts for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

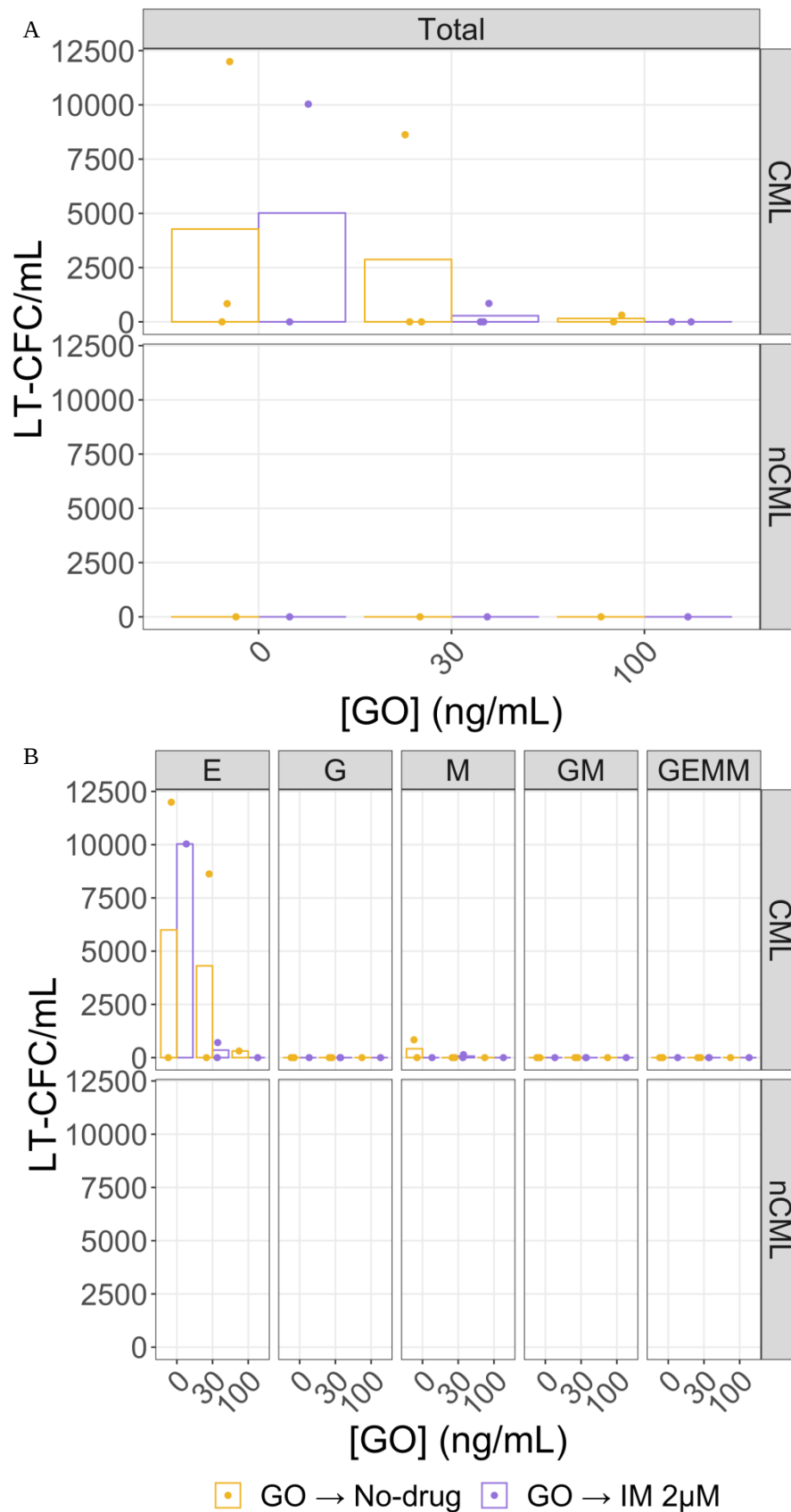


Figure 5-28. CFC concentration after sequential treatment GO → IM/no drug and 6 weeks on culture with stroma cells. (A) Summarised CFC concentrations for each treatment and condition. (B) CFC concentrations for each subtype. Individual dots are the colony counts for each patient and bars are the mean.

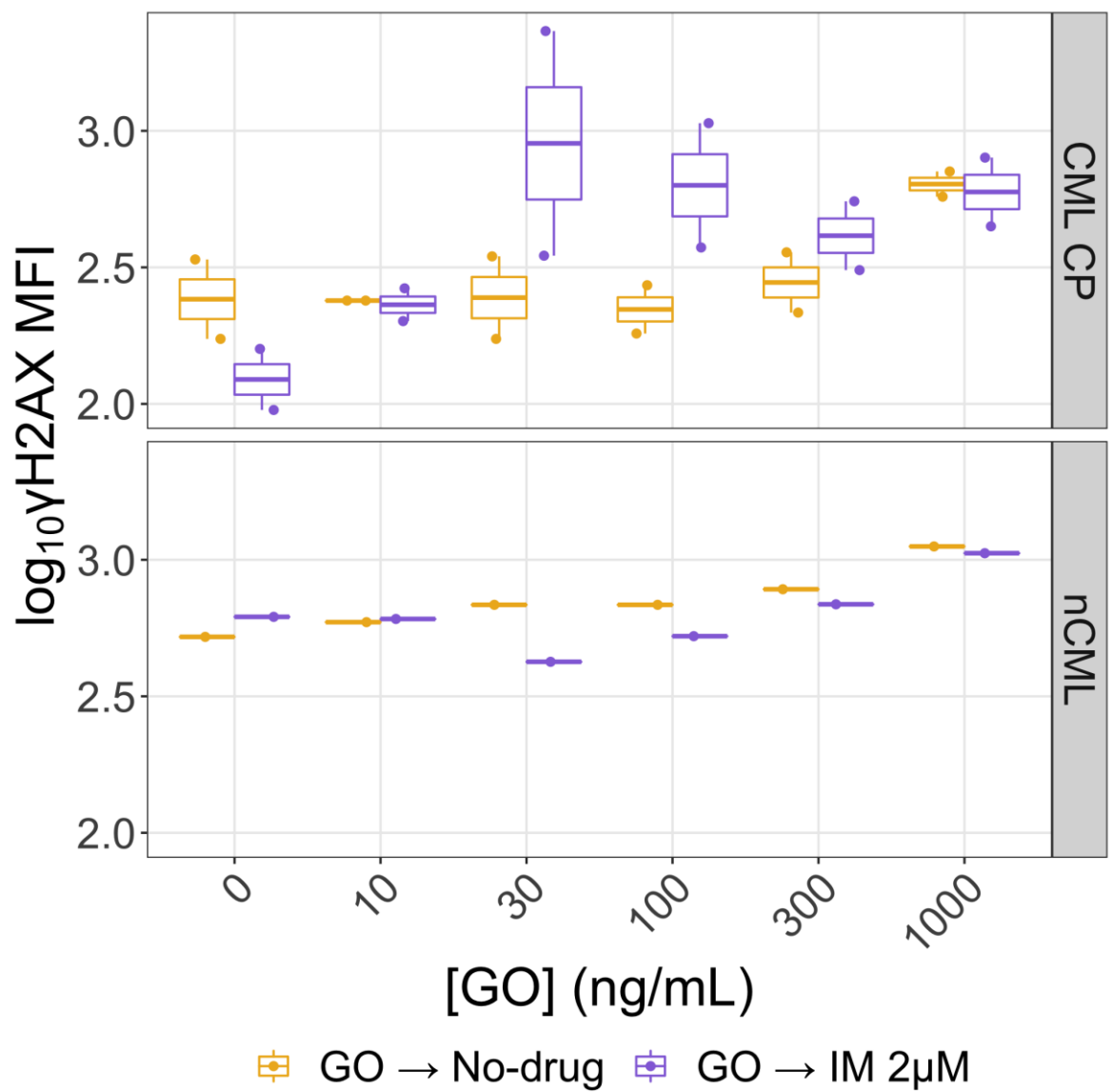


Figure 5-29. IM treatment after GO reduced the concentration of GO required for increasing the levels of γH2AX increases in CML. CML cells seem to increase the levels of γH2AX in response to increasing concentrations of GO, especially with concentrations between 100 and 1000ng/mL, and between 0 and 30ng/mL when the cells were treated with IM after GO. The nCML samples increase γH2AX with increasing GO concentrations but the correlation was not statistically significant. Central line in the boxes represents the median of the values. The dots represent the individual samples analysed in the experiment. Replicates: $n_{\text{CML}}=1$, $n_{\text{nCML}}=2$.

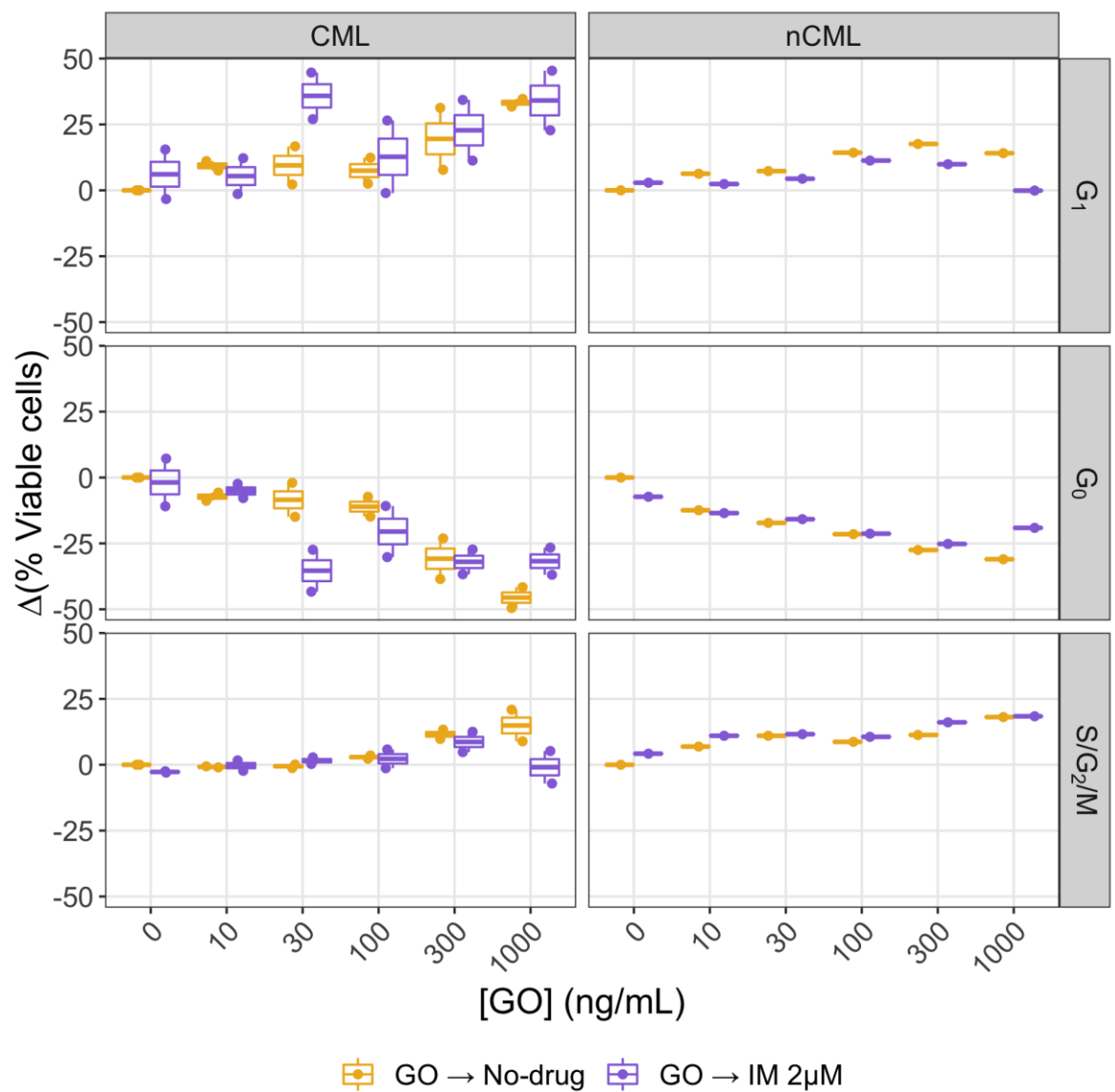


Figure 5-30. Differences on the percentage of cells in each cell cycle phase compared with the NDC after sequential treatment GO → IM/no drug. A trend towards a more proliferative phenotype can be observed in CML and nCML cells. The difference in cells in $S/G_2/M$ and G_1 phases positively correlated with the logarithm of GO concentration in CML and nCML while G_0 phase was found to have a lower percentage of cells with increasing concentrations of GO. Central line in the boxes represents the median of the values. The dots represent the individual samples analysed in the experiment. Replicates: $n_{\text{CML}}=2$, $n_{\text{nCML}}=1$.

5.2.28 Understanding the mechanism of action of GO by investigating whole transcriptome expression changes by RNAseq after 72h of treatment

The cell biology experiments on the use of GO as a potential treatment in CML revealed that it is effective at eradicating CML CD34⁺ cells. This was mediated by an increase in DNA double strand breaks, as reported by the increase of γ H2AX and it induces entry into the cell cycle. In order to further investigate the mechanism by which GO eliminates CML CD34⁺ cells, it was decided to investigate whole-transcriptome changes in an RNAseq experiment. CML cells were studied under the effect of IM 2 μ M, GO 100ng/mL and the combination of both drugs after 72h of treatment and compared with the NDC. The 72h treatment regimen was chosen for the RNAseq experiment because of the wider therapeutic window and under the assumption that the cells would have the lesser amount of culture-related changes. The samples sequenced were those from patients CML423, CML441 and CML460 (Table 2-1). No nCML samples were sequenced due to economical constraints, so CML samples were prioritised.

RNA was extracted using the Arcturus PicoPure kit for patients CML423 and CML460 while RNAeasy Micro-kit was used for the samples from CML441. All samples had RIN number above 8 with most of them over 9. Library preparation was performed by Glasgow Polyomics using SMART-seq ultra low cDNA conversion kit and Nextera library preparation kit. Sequencing was performed in the Illumina NextSeq and HiSeq machines (a previous run on HiSeq was complemented with a second run on NextSeq). The number of reads per samples was fairly constant, with most samples ranging between 40 and 50 million aligned reads and only two having more than 50 million (Figure 5-31A).

Initial exploration of data using PCA showed a clear separation between the different treatments as well as a separation between the patients CML423 and CML460 to CML441 (Figure 5-31B). IM treated cells were the closest ones to the NDC followed by the GO treated cells and finally the combination treatments were the farthest from the NDC. The proximity of the IM treated cells could be explained by the IM insensitivity that at least CML423 and CML441 presented in the clinical setting (no information available for CML460; Table 2-1). The volcano plots showed a trend towards smaller p-values in up-regulated genes compared with the downregulated ones (Figure 5-32 and Figure 5-33). As the PCA anticipated, IM was the treatment that induced the lowest number of gene expression changes (88 genes), followed by GO (257 genes) and the combination (5,280 genes). When the combination treatment was compared with only the IM treated cells 4,199 genes were differentially expressed. The total amount of genes mapped in the sequencing was 19,981.

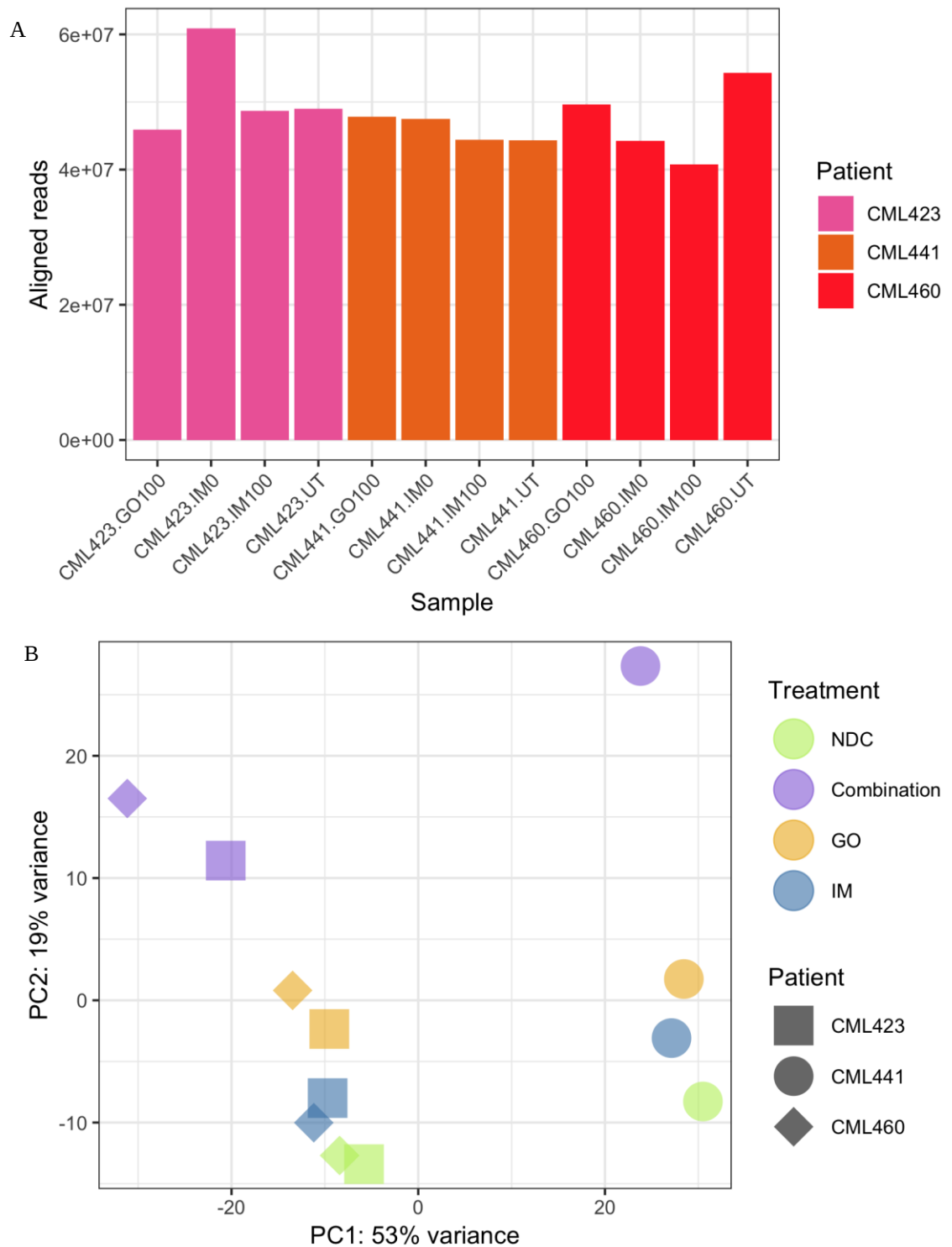


Figure 5-31. Initial exploration of the RNAseq data did not reveal unexpected sources of variability. (A) All the samples presented more than 40 million reads and only *CML423 IM0* (IM alone) and *CML460 UT* (NDC) had more than 50 million reads. (B) The PCA showed patient related clustering in PC1, with *CML441* forming a much separated cluster (different RNA extraction kit, age and sex). PC2 showed a clear separation of the different treatments, with IM being the closest to NDC and the combination treatment the farthest one.

The differentially expressed genes for each treatment were compared with the list of TKI₆₀ genes described in Chapter 3. Of the 60 TKI₆₀ genes originally described, 46 were present in the universe of the RNAseq analysis. It was found that GO treatment upregulated *EGFL6* and *ZMAT3* (Figure 5-34, Figure 5-35A), which were originally also upregulated in CML compared with nCML. However, an overlap of 2 TKI₆₀ genes with the genes differentially expressed after GO treatment was not found statistically significant after hypergeometric distribution ($p=0.16$) or Monte-Carlo ($p=0.14$). IM was also found to deregulate 2 TKI₆₀ genes ($p=0.025$), *C10orf10* (upregulated) and *CD33* (downregulated; Figure 5-34, Figure 5-35A). This did not agree with the other techniques used to assess if *CD33* gene expression is affected by IM in this thesis. The number of TKI₆₀ genes that were differentially expressed by the combination treatment compared with the NDC was 21 ($p_{\text{hypergeometric}}=0.042$, $p_{\text{Monte-Carlo}}=0.008$) and 15 when compared with the IM alone arm ($p_{\text{hypergeometric}}=0.187$, $p_{\text{Monte-Carlo}}=0.064$) (Figure 5-34, Figure 5-35B and C).

With the results of the RNAseq analysis it was decided to investigate if there was any pathway or level 3 GO-term overrepresented in the different lists of differentially expressed genes using CPDB (Kamburov et al., 2013). Changes in response to GO seemed to relate mainly with platelet activation pathways, inflammation and p53 (Table 5-1). The same could be observed from the level 3 GO-terms (Table 5-2). On the other hand, genes differentially expressed in response to IM seemed to be overrepresented in members of the GPCR pathways and platelet activation (Table 5-3) as were the GO-terms (Table 5-4). The combination treatment had an overrepresentation of members of GPCR, inflammation and platelet activation pathways when compared to NDC (Table 5-5) and lay towards inflammation and p53 signalling when compared to IM alone instead of the NDC (Table 5-7). The GO-terms for which there was an overrepresentation in the combination treatment were also related with immune and platelet activation, and regulatory processes (Table 5-6). When compared with the IM only treatment the combination was enriched for membrane components, extracellular vesicles and immune activation GO-terms (Table 5-8). The complete list of pathways and GO-terms can be found in Appendix II and Appendix III, respectively.

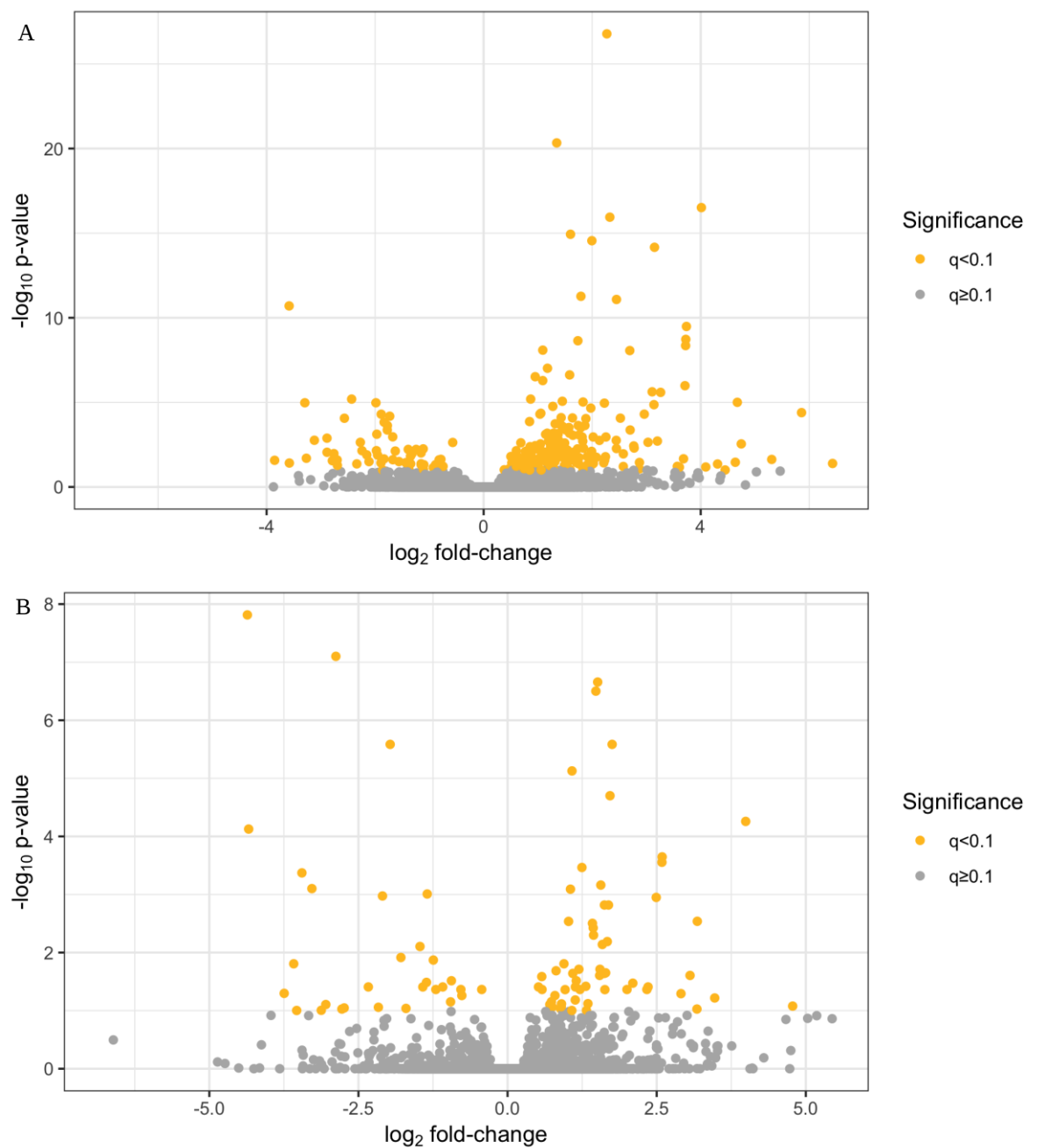


Figure 5-32. Volcano plots of the single treatment. (A) Volcano plot showing the changes in gene expression after 72h of GO 100ng/mL. (B) Gene expression changes after 72h IM 2 μ M. The statistically significant differentially expressed genes are coloured in orange.

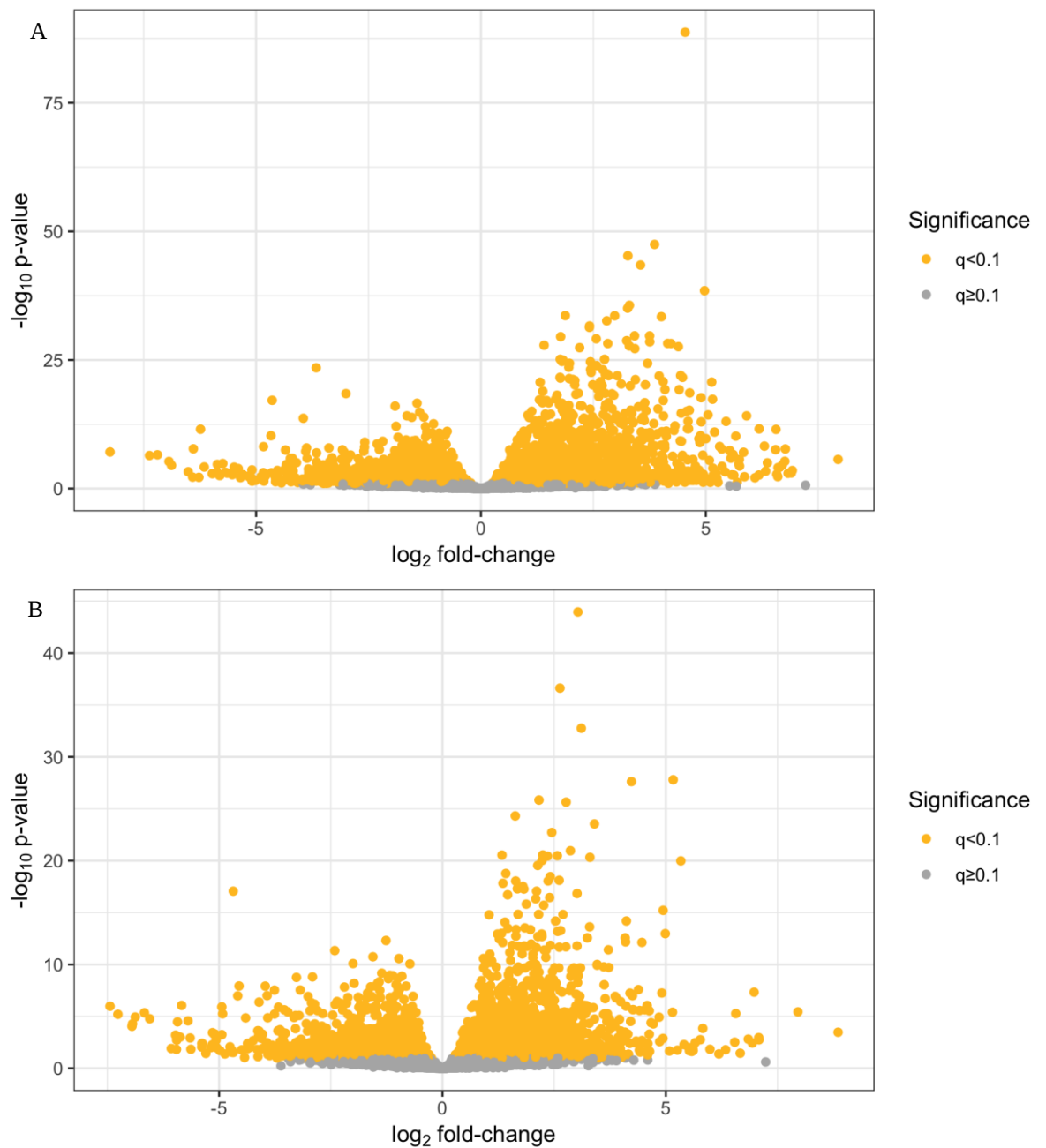


Figure 5-33. Volcano plots showing the gene expression changes after the combination treatment. (A) Compared with the NDC. (B) Compared with 72h IM 2 μ M. The statistically significant differentially expressed genes are coloured in orange.

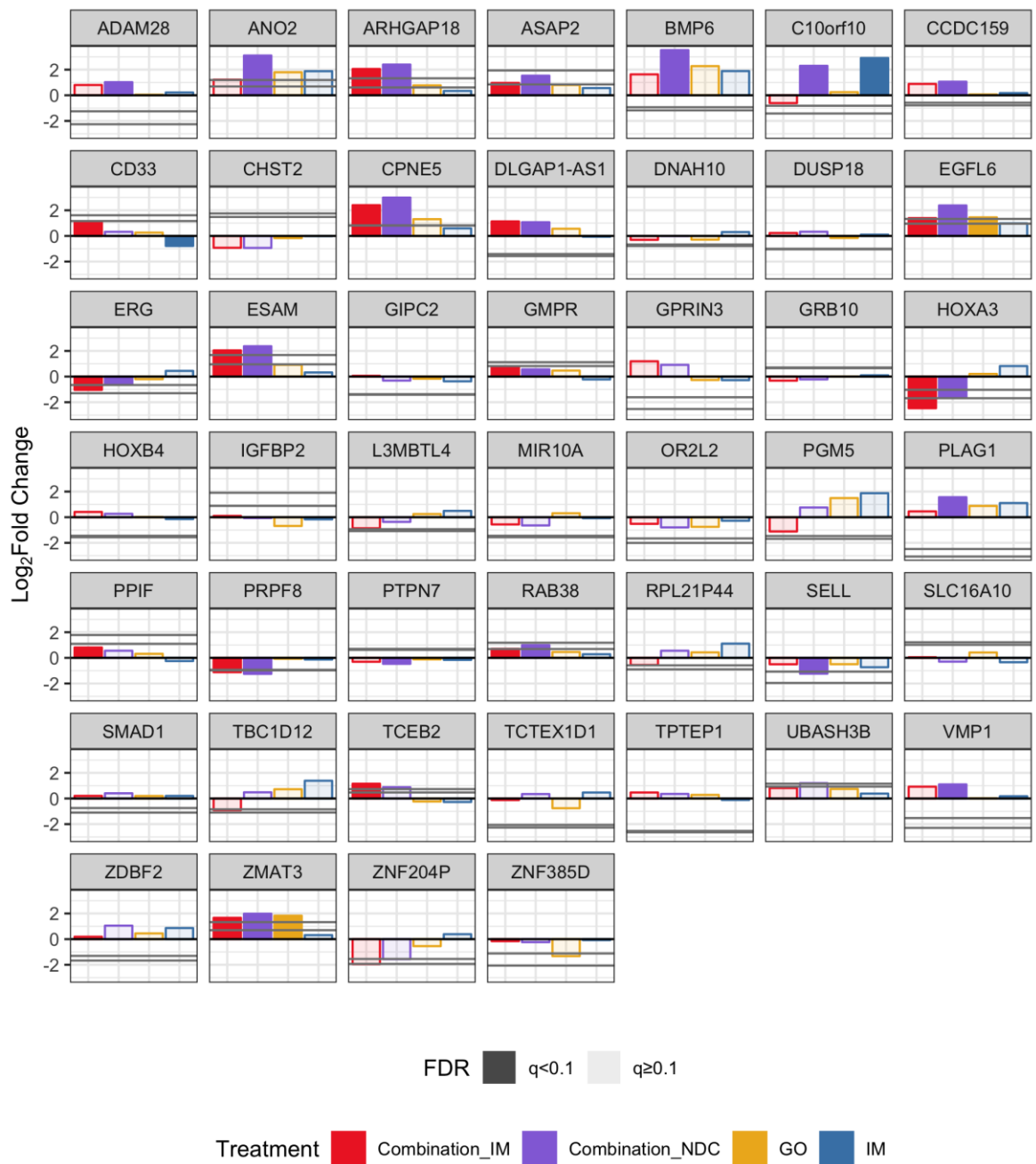


Figure 5-34. Gene expression changes after the different treatments on the TKI₆₀ genes present in the analysis. The black horizontal line marks the 0 and the grey horizontal lines mark the log₂Fold Changes of the CMLDV and CMLMC datasets. Filled bars are those with a q-value smaller than 0.1.

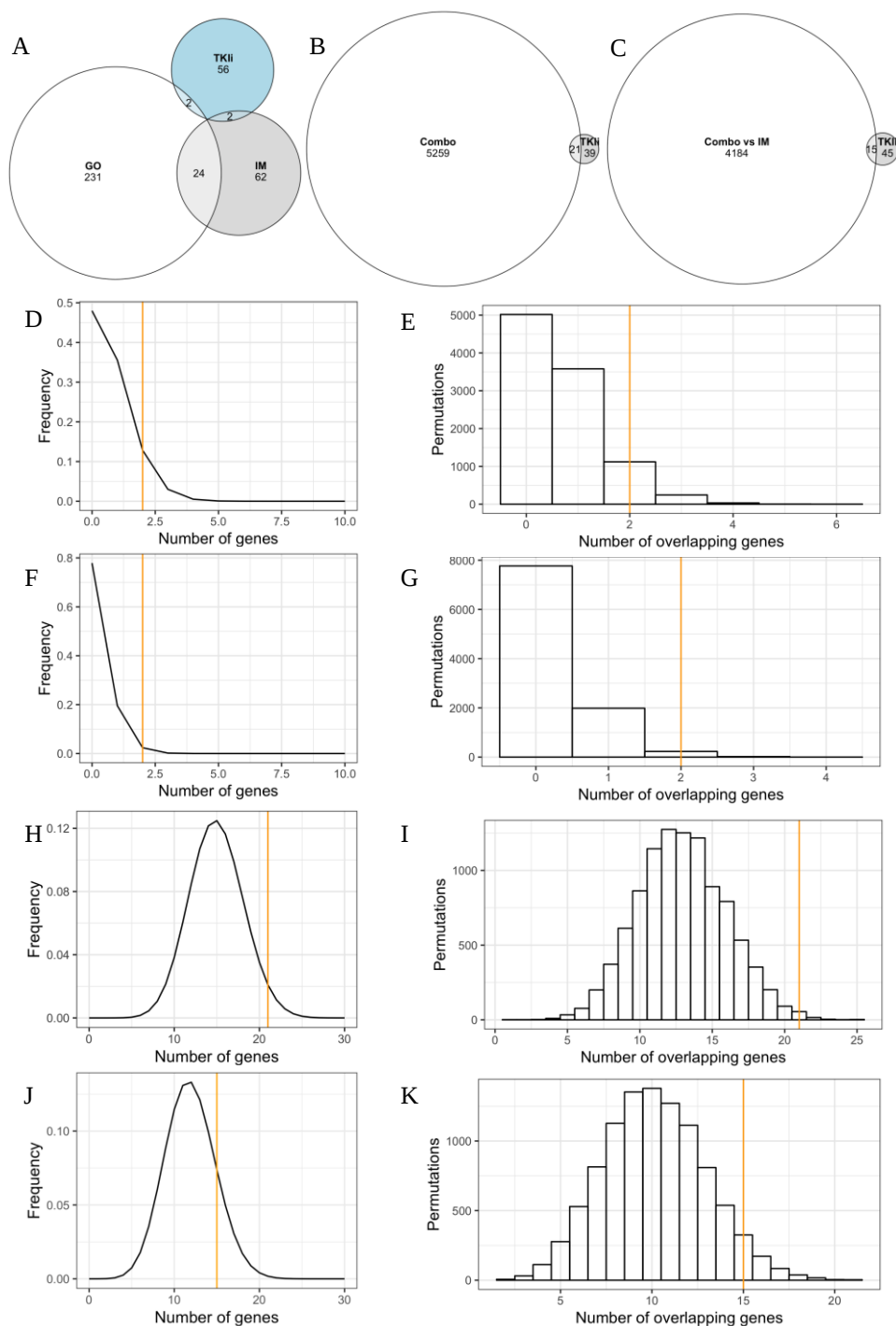


Figure 5-35. Overlaps between the different treatments and the TKI₆₀ genes. (A) Overlap between TKI₆₀ genes and the genes differentially expressed after GO and IM treatments. (B) Overlap between the genes differentially expressed after the combination treatment and the TKI₆₀ genes. (C) Overlap between the genes differentially expressed after the combination treatment when compared with the IM only treatment and the TKI₆₀ genes. (D) Hypergeometric distribution and (E) Monte-Carlo permutations for the probability of an overlap between the GO differentially expressed genes and the TKI₆₀ genes. (F) Hypergeometric distribution and (G) Monte-Carlo for the overlap between the IM differentially expressed genes in the RNAseq and the TKI₆₀ genes. (H) Hypergeometric distribution and (I) Monte-Carlo for the overlap between the combination treatment differentially expressed genes and the TKI₆₀ genes. (J) Hypergeometric distribution and (K) Monte-Carlo for the overlap between the combination treatment differentially express genes when compared with IM alone and the TKI₆₀ genes. The orange lines represent the true overlap between the two lists compared in each figure.

Table 5-1. Top 10 overrepresented pathways in the genes differentially expressed after 72h of treatment with GO, as calculated by CPDB.

Pathway	Source	q-value
Haemostasis	Reactome	4.78E-12
Platelet activation, signalling and aggregation	Reactome	1.17E-09
p53 signalling pathway - Homo sapiens (human)	KEGG	2.99E-09
Platelet degranulation	Reactome	3.31E-09
Response to elevated platelet cytosolic Ca ²⁺	Reactome	4.79E-09
Formation of Fibrin Clot (Clotting Cascade)	Reactome	1.91E-08
Cell surface interactions at the vascular wall	Reactome	8.96E-07
Cytokine-cytokine receptor interaction - Homo sapiens (human)	KEGG	2.56E-06
Common Pathway of Fibrin Clot Formation	Reactome	3.98E-06
Complement and coagulation cascades - Homo sapiens (human)	KEGG	7.46E-06

Table 5-2. Top 10 overrepresented GO-terms in the genes differentially expressed after 72h of treatment with GO, as calculated by CPDB.

Name	ID	q-value
blood coagulation	GO:0007596	9.27E-11
response to wounding	GO:0009611	2.41E-09
regulation of body fluid levels	GO:0050878	5.29E-09
platelet activation	GO:0030168	5.29E-09
regulation of signalling	GO:0023051	6.75E-07
regulation of response to stimulus	GO:0048583	6.75E-07
cell surface receptor signalling pathway	GO:0007166	7.31E-07
positive regulation of biological process	GO:0048518	7.31E-07
regulation of haemostasis	GO:1900046	8.50E-07
exocytosis	GO:0006887	1.57E-06

Table 5-3. Overrepresented pathways in the genes differentially expressed after 72h of treatment with IM, as calculated by CPDB.

Pathway	Source	q-value
Complement and coagulation cascades - Homo sapiens (human)	KEGG	0.00963555
GPCR ligand binding	Reactome	0.00963555
Formation of Fibrin Clot (Clotting Cascade)	Reactome	0.00963555
Signalling by GPCR	Reactome	0.02102657
Class A/1 (Rhodopsin-like receptors)	Reactome	0.03701184

Table 5-4. Top 10 overrepresented GO-terms in the genes differentially expressed after 72h of treatment with IM, as calculated by CPDB.

Name	ID	q-value
regulation of multicellular organismal process	GO:0051239	3.30E-06
regulation of response to stimulus	GO:0048583	5.62E-05
cell surface receptor signalling pathway	GO:0007166	5.62E-05
response to organic substance	GO:0010033	0.00053477
negative regulation of biological process	GO:0048519	0.00054523
regulation of haemostasis	GO:1900046	0.00057476
regulation of signalling	GO:0023051	0.0014484
blood coagulation, fibrin clot formation	GO:0072378	0.0018736
regulation of developmental process	GO:0050793	0.00228845
positive regulation of biological process	GO:0048518	0.0031739

Table 5-5. Top 10 overrepresented pathways in the genes differentially expressed after the combination treatment compared with the NDC, as calculated by CPDB.

Pathway	Source	q-value
Haemostasis	Reactome	9.41E-07
Response to elevated platelet cytosolic Ca ²⁺	Reactome	7.01E-06
Platelet degranulation	Reactome	5.95E-05
Cytokine-cytokine receptor interaction - Homo sapiens (human)	KEGG	0.00015908
Platelet activation, signalling and aggregation	Reactome	0.00015908
Chemokine receptors bind chemokines	Reactome	0.00103075
Nucleotide-like (purinergic) receptors	Reactome	0.00415011
G alpha (i) signalling events	Reactome	0.01110047
Complement and coagulation cascades - Homo sapiens (human)	KEGG	0.01322177
RHO GTPases activate PKNs	Reactome	0.01322177

Table 5-6. Top 10 overrepresented GO-terms in the genes differentially expressed after the combination treatment compared with the NDC, as calculated by CPDB.

Name	ID	q-value
cellular response to chemical stimulus	GO:0070887	1.50E-05
regulation of immune system process	GO:0002682	1.50E-05
response to wounding	GO:0009611	1.67E-05
response to oxygen-containing compound	GO:1901700	1.67E-05
response to organic substance	GO:0010033	1.67E-05
negative regulation of biological process	GO:0048519	1.67E-05
extracellular vesicle	GO:1903561	3.76E-05
regulation of localization	GO:0032879	1.75E-05
regulation of response to stimulus	GO:0048583	2.24E-05
regulation of multicellular organismal process	GO:0051239	4.15E-05

Table 5-7. Top 10 overrepresented pathways in the genes differentially expressed after the combination treatment compared with IM treated cells, as calculated by CPDB.

Pathway	Source	q-value
Cytokine-cytokine receptor interaction - Homo sapiens (human)	KEGG	3.25E-07
Response to elevated platelet cytosolic Ca ²⁺	Reactome	5.74E-07
Haemostasis	Reactome	1.56E-06
Platelet degranulation	Reactome	1.88E-06
Chemokine receptors bind chemokines	Reactome	1.35E-05
Platelet activation, signalling and aggregation	Reactome	1.77E-05
p53 signalling pathway - Homo sapiens (human)	KEGG	0.00012155
Neutrophil degranulation	Reactome	0.00015719
GPCR ligand binding	Reactome	0.00135731
G alpha (i) signalling events	Reactome	0.00140623

Table 5-8. Top 10 overrepresented GO-terms in the genes differentially expressed after the combination treatment compared with IM treated cells, as calculated by CPDB.

Name	ID	q-value
extracellular vesicle	GO:1903561	1.44E-08
intrinsic component of plasma membrane	GO:0031226	6.81E-08
integral component of membrane	GO:0016021	7.92E-07
exocytosis	GO:0006887	8.14E-06
cell activation involved in immune response	GO:0002263	8.14E-06
leukocyte mediated immunity	GO:0002443	1.14E-05
regulation of response to stimulus	GO:0048583	1.78E-05
regulation of immune system process	GO:0002682	1.78E-05
leukocyte degranulation	GO:0043299	3.21E-05
chemotaxis	GO:0006935	3.68E-05

5.3 Discussion

This chapter shows the benefits of using GO for targeting CML cells *in vitro*. In all three treatment regimens, GO showed a lower IC₅₀ in CML than in nCML cells, suggesting that its use would be relatively safe for non-leukaemic cells. This is also supported by the data collected from clinical trials showing its safety even at higher concentrations than the ones used in this chapter (Castaigne et al., 2012). Additionally, the combination of GO and IM seems to have an additive like effect (with small changes in the effect when compared with the expected effect), which suggest that both drugs could be used together for the treatment of CML. These results are also supported by a previous report in the use of GO in the treatment of CML (Herrmann et al., 2012). They showed a synergetic effect of GO and TKIs (nilotinib and bosutinib) on treating CML mononuclear cells. However, they did not study the effect on LSCs or normal controls and did not investigate the mechanism by which GO target CML cells. Also, the culture conditions did not approximate physiological conditions and did not investigated the effect of GO in cells previously treated with IM. Therefore, the results presented in this chapter add novelty to the field.

IM has been shown to significantly affect a larger number of TKI₆₀ genes after 72h of treatment at 2µM than expected by chance. However, the small amount of genes affected (two: *C10orf10* and *CD33*) means that a change of one would mean a non-significant number of TKI₆₀ affected genes and the results should be interpreted carefully. Although their expression had been shown to not be affected by IM treatment in Chapter 3 both in the microarray analysis (Figure 3-6) and by qPCR (Figure 3-11), culture conditions differed (7 days with no growth factor in Chapter 3 and 3 days in physiological conditions in the current chapter), which could explain the observed differences. However, the *CD33* microarray and qPCR results were supported in this chapter at the protein level when the amount of CD33 on the cell surface was measured by flow cytometry (Figure 5-6). However, the RNAseq analysis showed a modest decrease in the levels of *CD33* expression after IM treatment. This can be caused by a technique specific artefact. As mentioned in the introduction of this chapter, CD33-related Siglecs have very similar sequences within the same species (Crocker et al., 2007) and this could have affected the quality of the alignment and transcript identification. In any case, the protein levels on the cell's surface remain constant after IM treatment and the protein is the ultimate target of GO, so the ability for binding CML cells by GO should not be compromised by IM treatment.

One of the main pitfalls of TKI treatment is its inability to target quiescent CML cells, which accumulate during TKI treatment (Graham et al., 2002, Copland et al., 2006,

Jorgensen et al., 2007, Giustacchini et al., 2017). This happens due to the anti-proliferative effect of TKIs on CML LSCs, which allows them to accumulate in G₀ and survive. Previous attempts at inducing CML LSCs to enter cell cycle using G-CSF increased the sensitivity to IM treatment of CML cells (Jorgensen et al., 2006) and it was shown to be safe in clinical practice (Drummond et al., 2009). In this chapter it has been shown that GO is able to induce CML (and nCML) cells into cell cycle entry even in the presence of IM, which could lead to the eventual exhaustion of the leukaemic clone and the eradication of the disease. This is also supported by the CFC and LTC-IC assay results, especially those from the GO→IM regimen. Although not statistically significant, the magnitude of the effect shows the benefit of this combination for reducing the number of CFCs in CML.

The increase in the levels of γ H2AX in GO-treated cells already pointed to GO producing double strand breaks in the DNA of treated cells, as its proposed mechanism of action suggested. However, it was thought that a more detailed study of the mechanism by which GO targets CML cells and how this compared to IM could be beneficial.

The reduced number of detected differentially expressed genes by RNAseq in the single drug treatments was unexpected, especially for IM. However, at least two of the patients are known to not respond well to IM treatment in the clinical setting: one because of reported intolerance to IM and NIL and another responded well to BOS after 3 months on IM. Therefore, it is possible that the poor response of these patients to the treatment in the clinics is translated into a milder transcriptomic response *in vitro*. However, it is also possible that 72h of treatment *in vitro* is not enough time for the transcriptomic changes induced by IM to be detected, especially compared with more drastic conditions, such as the drug combination treatment. After GO treatment a modest number of genes were found differentially expressed, although more changes were observed in response to GO than to IM. On the other hand, the combination treatment did have a big effect in gene expression, with statistically significant changes in about 25% of the analysed genes. The top 10 most overrepresented pathways were related to platelet function, inflammation and/or G-protein coupled receptor (GPCR) pathways. GPCRs are involved in platelet activation and inflammation response, so all three categories might be considered the same. These pathways might be overrepresented in response to cell death due to drug treatment and might be nonspecific. However, this could also happen because of the activation of LSCs and the generation of progenitors as a repopulation response, as the increase in cycling cells might suggest.

The genes differentially expressed after GO treatment were overrepresented in genes belonging to the p53 signalling pathway. An increase in the expression of *ZMAT3* and its negative feedback regulator *MDM2* was observed (Xirodimas et al., 2004, Prives, 1998). The p53 pathway is known to be activated in response to DNA damage and it activates apoptosis if the cell is unable to recover from the damage. This might suggest an activation of p53 pathway in response to DNA damage induced by GO.

Treatment with the combination treatment affects the expression of 21 out of the 46 TKI_{i60} genes included in the analysis (i.e. genes with more than 100 reads). Direction of the gene expression was opposite to that observed in the CML vs normal comparison in chapter 3 in genes such as *BMP6* and *PLAG1*. *PLAG1* is known to be upregulated in a subset of AML patients and it has been found to interact with CBF, blocking differentiation and inducing proliferation (Landrette et al., 2005). While this gene was downregulated in CML compared with normal samples it is upregulated in response to the combination treatment, suggesting that it participates in the increasing percentage of cells entering cell cycle. BMP family proteins have also been found upregulated in TKI-resistant patients receiving TKI treatment (Grochowski et al., 2017, Toofan et al., 2018) and therefore, this could be related with the fact that the patients used in this experiment did not respond well to IM treatment in the clinic. Furthermore, BMP family proteins have been found to be important in the survival of CML LSCs and the dual treatment with BMP inhibitors and TKIs impairs self-renewal in those cells (Toofan et al., 2018). Therefore, its upregulation of *BMP6* in treated samples may constitute a survival factor.

Other genes like *ERG* or *HOXA3* are deregulated in the same direction than they were in the comparison between CML and normal cells. *ERG* is known to regulate HSC differentiation via repression of *MYC* and it promotes quiescence (Knudsen et al., 2015). The downregulation phenotype observed in the combination treatment strengthens the reduction of the percentage of cells in G₀ observed in the cell cycle analysis. HOX genes have also been found to be important for the maintenance of stem cell phenotype in HSCs (Sauvageau et al., 1994) and low levels of HOX genes expression have been found to be a good prognosis marker in AML (Gollner et al., 2017). In this analysis *HOXA3* was found to be downregulated after the combination treatment, potentially favouring proliferation and cell differentiation.

The results from the GO→IM regimen suggest that GO has medium-term effects on CML CD34⁺ cells at concentrations of 100ng/mL or higher that last at least 72h after removal of the drug. This can be considered at the time of designing a treatment regimen as lower

concentrations will probably have little therapeutic value in a medicine that is delivered every two weeks as GO (EMA, 2008). Additionally, during this treatment regimen, the GO+IM combination presented a moderate synergy (6.87%), which was not observed in any of the other treatment regimens (the simultaneous combination was even slightly antagonistic). This can be clearly appreciated in the effect that the treatment has in the cell cycle. While the decrease of CML cells on G_0 happens at lower concentrations of GO in the GO→IM regiment (Figure 5-30), in the simultaneous combination IM seems to antagonise the effect of GO and moderated the reduction of CML cells in G_0 (Figure 5-14). It is also worth noticing the refractory effect that IM has in the cell cycle. When CML cells were treated with IM and then IM was washed away, it is possible to observe a decrease of CML cells in G_0 in the cells that recovered for 72h when compared with the NDC (Figure 5-22).

Treatment with GO seems to promote cell cycle entry in parallel to the cell death induced by DNA damage. The fact that cell cycle entry is induced even in the presence of IM, which is known to increase the proportion of quiescent cells (Graham et al., 2002), and the observed gene expression changes suggests that GO may induce cell proliferation and differentiation. Additionally, its combination with IM has additive or close to additive responses for cell counts, meaning that its use should not have any negative effect on the efficacy of the TKI treatment. Thus, these results suggest that the use of concentrations of GO of no lower than 100ng/mL in combination with TKI treatment could be beneficial for the treatment of CML. However, these experiments should be expanded into a larger cohort of patients' primary cells and complemented with *in vivo* experiments to confirm the safety of the combination treatment before translation into a clinical trial.

6 General discussion

6.1 New therapeutic approaches are necessary for curing CML

CML is currently a well-managed disease with expected life-span very close to those of the general healthy population. However, current TKI treatments are unable to eradicate the LSCs and the leukaemic clone persists in the patients. This is a psychological burden for CML patients. In addition, the increasing prevalence of CML and the high cost of the treatments remain a challenge to public health services, such as the NHS in the UK and the SNS in Spain, and it affects the availability and compliance of the treatment in countries without public health services or with lower funding (Abboud et al., 2013, Beinortas et al., 2016, Kurtovic-Kozaric et al., 2016). However, the economic burden of CML treatment has decreased since IM became available off patent (Lejniec et al., 2017). Although some studies have shown the possibility of safely discontinuing TKI treatment in 4% of CML patients this is not effective in all patients (Chomel et al., 2016, Mahon et al., 2010). Thus, new therapeutic approaches that target the LSCs and cure the disease are needed in CML.

The effect of TKIs in CML LSCs has been well studied. First, IM was shown to have a limited effect in the quiescent CD34⁺ cell population while it eradicates the cycling cells (Graham et al., 2002). Moreover, IM seemed to have an antiproliferative effect on the cells as a higher number of quiescent CML CD34⁺ cells were found when treated with IM than in the NDC. A similar effect was described for NIL (Jorgensen et al., 2007). DAS was found to have a more potent inhibition of the TK activity of BCR-ABL1 in CML LSCs than IM but it was found to enrich the CML cells in the quiescent LSCs population (Copland et al., 2006). A recent sequencing report (Giustacchini et al., 2017) confirmed that a CML LSC quiescent population present at diagnosis (before TKI treatment), of primitive HSC phenotype, was enriched in the patients' blood and bone marrow over time when treated with TKI. Following this evidence, other reports (Charaf et al., 2016, Zhang et al., 2018) showed that IM may not be enriching for quiescent CML LSCs just by eliminating the cycling cells but by inducing the expression of quiescence and self-renewal genes in both CML and normal cells. This shows that TKIs induce a phenotype in CML cells that protects them from the treatment. Therefore, additional treatments other than TKI are required for complete eradication of the leukaemic clone.

The antiproliferative effect of TKIs and the induction of a stem/quiescent phenotype is potentially the leading cause of TKIs failure in eradicating CML cells. Therefore, induction of cell-cycle entry in those cells could potentially re-sensitise CML TKI persistent cells to TKI treatment. The use of interferon- α (IFN α) has been proposed as a

potential agent to use in combination with TKI because of its effect in inducing cell-cycle entry in murine primitive HSCs (Essers et al., 2009). Furthermore, intermittent doses of G-CSF induce cell-cycle entry in quiescent human CML CD34⁺ cells and increase the sensitivity of CML cells to IM *in vitro* (Jorgensen et al., 2006). However, although the combination of pulsing IM and pulsing G-CSF seemed non-toxic in a phase II clinical trial, a higher number of patients lost CCyR or MMR in the study group than the group taking the standard 400mg/day IM treatment (Drummond et al., 2009), suggesting that the benefits of the pulsing G-CSF do not compensate for the interrupted treatment with IM.

The persistence of CML LSCs during TKI treatment opened the question of what enables this population to survive treatment at the same time retaining leukaemic properties. The finding that it is possible to rescue CML progenitor cells (CD34⁺CD38⁺) from specific inhibition of the BCR-ABL1 TK activity by treatment with stem cell factor (SCF), a KIT ligand, suggested that similar mechanisms independent of the BCR-ABL1 TK could be involved in the persistence of CML LSCs to TKI treatment (Corbin et al., 2013). Similarly, β -catenin pathway was shown to be upregulated in CML LSCs that persist TKI treatment compared with those that are sensitive to it, and that this increase in β -catenin was independent from extracellular (stroma) signalling (Eiring et al., 2015). These findings supported a previous report showing that dual BCR-ABL1 inhibition by knock-down and dasatinib treatment reduced the phosphorylation of BCR-ABL1 TK targets such as STAT5 and CRKL but failed to completely eradicate the leukaemic clone, which was enriched in primitive LTC-IC (Hamilton et al., 2012). BCR-ABL1 TK independent gene expression de-regulation has previously been observed in CML in a miRNA screening that reported *MIR10A* to be downregulated in CML even during TKI treatment, conferring a growth advantage to the leukaemic cell (Agirre et al., 2008). This suggested that there is a BCR-ABL1 TK independent signature in CML that allows the leukaemic clone to survive under inhibition of BCR-ABL1 TK by TKIs while retaining the leukaemic properties.

These results highlighted the need to find new therapeutic targets in CML other than the TK of BCR-ABL1. As TKIs are effective at eradicating cycling cells, most efforts have focused in targeting elements that promote stemness and quiescence in the CML LSCs. The high level of JAK2 activity observed in CML LSCs even during BCR-ABL1 TK inhibition led to the use of PP2A activating drugs (PADs). PADs were successful at inhibiting JAK2 pathway's activity and reduced the number of CML LSCs but failed to completely eradicate the leukaemic clone (Neviani et al., 2013). Similarly, although STAT5 activity is reduced after TKI treatment in CML LSCs, the gene expression and protein levels remain constant. The use of pioglitazone, an agonist of PPAR γ , was shown

to be effective at reducing the levels of STAT5 expression and, therefore, its residual activity (Prost et al., 2015). This reduction of STAT5 activity did in fact induce cell cycle entry and reduced the colony forming capacity of the leukaemic clone. Additionally, it was shown to be effective at inducing complete molecular response in 3 CML patients that were already on IM treatment (Prost et al., 2015). Additionally, the deregulation of Polycomb Repressive Complex 2 (PRC2) targets in CML and both the downregulation of *EZH1* and the upregulation of *EZH2* led to the use of EZH2 inhibitors (EZH2i) for targeting CML LSCs (Scott et al., 2016). EZH2i treatment seemed to potentiate the transcriptional signature of TKI on CML LSCs and, despite not affecting normal haematopoiesis, it induced apoptosis in CML LSCs when combined with TKIs (Scott et al., 2016). Autophagy has been shown to be a protective mechanism in cancer cells against anti-cancer therapy. The use of autophagy inhibitors in CML has shown a reduced number of CML LSCs, which enter cell cycle and differentiate and, therefore, are targeted by TKI (Baquero et al., 2018). Finally, a deregulated signalling network with MYC and p53 as main nodes has been described in CML LSCs (Abraham et al., 2016). Targeting this network using both MDM2 inhibitors (MDM2i) and bromodomain and extra terminal protein inhibitors (BETi) for stabilising p53 and inhibiting MYC transcriptional activity, respectively, is able to target genes not affected by TKIs and induces both apoptosis and differentiation, opening a new avenue for CML therapy (Abraham et al., 2016).

Although some of these approaches target genes and/or pathways not affected by TKI treatment, a BCR-ABL1 TK-independent signature has not been defined before. TKIs are known to inhibit the TK of BCR-ABL1 (Hamilton et al., 2012). Therefore, the use of TKI treated CML CD34⁺ cells for investigating the genes independent of BCR-ABL1 TK was considered to be the closest approximation to a naturally occurring CML cell with a non-active BCR-ABL1 TK despite the off-target effects. The use of a knock-down model would also eliminate the effect of the other domains of BCR-ABL1, such as its scaffolding activity, which is important for the role of JAK2 in CML LSCs (Neviani et al., 2013). Also, the vector integration would potentially affect the overall gene expression of the cells. Therefore, the TKI-independent (TKIi₆₀) genes were considered to be the best approximation to a BCR-ABL1 TK-independent genes list. Additionally, investigating the TKIi₆₀ genes response to the clinically relevant question of which CML de-regulated genes are not affected by the current TKI treatment.

The 60 genes TKIi₆₀ signature presented in the chapter 3 is then the first attempt to uncover the BCR-ABL1 TK independent signature in CML LSCs using whole transcriptome analysis. Although this does not reduce the validity of previous studies,

which have demonstrated their value by eliminating CML LSCs, it does add a rational for the development of new therapies that target the signature. In fact, the small number of patient samples used for uncovering the TKIi signatures is not enough to assume its presence in every CML patient. A larger sample size would have provided more statistical power, which would have increased the trust on the results by both reducing error type I and type II (Button et al., 2013, Sham and Purcell, 2014). Due to the limited availability of patients' material it was not possible to increase the power of the experiments. However, when possible, the results were cross-validated in more than one experiment (e.g. microarray and qPCR) and compared with the existing literature. For example, the TKIi₆₀ signature contains *MIR10A*, which have already been described as a TKI insensitive gene in CML (Agirre et al., 2008). Similarly, upregulation of the intrinsic β -catenin pathway has been previously reported, which could explain the low expression levels of *CDH2* (N-cadherin) in the microarray analysis. N-cadherin is a known activator of the extrinsic β -catenin pathway, which seems to have no effect on the persistence of CML LSCs to TKI treatment (Eiring et al., 2015). Additionally, PPIF is a known inhibitor of apoptosis which acts in a BCL2-like manner (Eliseev et al., 2009) and it has shown to directly interact with p53 in RAS induced tumours, reducing the antiproliferative effect of p53 in those cells (Bigi et al., 2016). Therefore, PPIF upregulation could contribute to the decrease in p53 pathway proteins reported previously (Abraham et al., 2016). *ERG* has been shown to be important in the transcriptional regulation of genes with GATA/RUNX and GATA/MYC motifs, activating the gene expression of the first and repressing the expression of the second group (Knudsen et al., 2015). Expression of *ERG* has shown necessary in normal HSCs for HSC maintenance and self-renewal and its down-regulation leads to HSC differentiation and exhaustion through over-expression of MYC regulated genes while expression of MYC is unaffected (Knudsen et al., 2015). Downregulation of *ERG* was rescued using BETi, recovering HSC self-renewal and decreasing differentiation (Knudsen et al., 2015). Interestingly, *ERG* activity was not found necessary for LSC maintenance in AML (Knudsen et al., 2015). Taken together, downregulation of *ERG* in CML LSCs might explain the deregulation of MYC pathway previously reported as well as the positive response to BETi treatment (Abraham et al., 2016). A graphical summary of this can be found in Figure 6-1.

Targeting the TKIi signature in CML LSCs has already shown promising results at eradicating CML LSCs (Abraham et al., 2016, Eiring et al., 2015). However, as part of this thesis it was decided to target directly one of the TKIi₄ genes, in contrast to targeting the pathways in which they are involved. The availability of a medicine for targeting CD33,

gemtuzumab ozogamicin (GO or MylotargTM) (Naito et al., 2000), made targeting CD33 a therapy that could rapidly be translated into the clinic. As shown in Chapter 5, targeting CD33 with GO in CML LSCs has a bigger effect than targeting normal HSCs, allowing for a therapeutic window in CML. Additionally, GO seems to induce cell cycle entry of CML LSCs and reduce colony forming potential, even when used in combination with IM, suggesting that it could contribute to the exhaustion of the leukaemic clone. However, this hypothesis should be tested in serial transplant experiments, which are considered in our material transfer agreement (MTA) with Pfizer. Although GO was already tested in CML (Herrmann et al., 2012), the work presented in this thesis was performed in a more primitive population, compares the effect with BCR-ABL1⁻ (nCML) controls and provides a better understanding of the mechanism of the treatment (cell cycle, DNA damage and general gene expression changes). Although TKIs have been shown safe for normal HSCs, it has been reported that inhibition of ABL1 by TKIs in normal cells can lead to an increased sensitivity to DNA damaging agents (Fanta et al., 2008). Therefore, a close monitoring of normal HSCs in long-term GO+TKI combination treatment would be needed to ensure that no additional mutations are induced in the normal HSC population.

Additionally, as shown in Chapter 4, the global direction of change of the TKI_{i60} genes between CML and normal controls is constant for most of the genes between the different phases of CML. However, because of the low sample size of the blast crisis group in CMLMC dataset this should be confirmed in an additional dataset, potentially Giustacchini (Giustacchini et al., 2017). If the results are confirmed, it would suggest that any treatment targeting the TKI_{i60} signature would potentially be also effective and safe for its use in the treatment of CML blast crisis.

Taken together, these results and the existing literature show that CML, characterised by the presence of a unique oncogene, *BCR-ABL1*, possesses a complex molecular signature that goes beyond the de-regulated activity of the TK of BCR-ABL1. Despite the success of managing the disease by the inhibition of this unique protein domain with TKIs, the complexity of CML LSCs prevents current approaches to eradicate the disease. As discussed here, new approaches, some of them targeting the TKI signature, have shown different success at eliminating CML LSCs. However, not a single approach has managed to completely eradicate this leukaemic population. Therefore, multiple pathways may need to be targeted simultaneous or sequentially in order to successfully eliminate the disease.

6.2 Biomarkers for TKI response and prognosis

Risk assessment of CML patients uses scores based in different clinical factors, such as age, percentage of blasts, spleen size and percentage of basophils (Sokal et al., 1984, Hasford et al., 1998, Hasford et al., 2011). Of note, the Sokal score (Sokal et al., 1984) was developed using standard (cytotoxic) chemotherapy treated patients survival and the Euro score (Hasford et al., 1998) using IFN γ treated patients. The newer EUTOS score (Hasford et al., 2011) is based in survival data of patients treated with IM in different European countries and uses the percentage of basophils and the spleen size at time of diagnosis for the assessment. However, despite the high specificity of the score predicting failure to achieve CCyR and progression free survival, the sensitivity is low (23% and 16% respectively). Therefore, patients classified as low-risk may not respond to TKI treatment and progress to a more aggressive phase of the disease. In order to improve patient classification and assess choice of treatment, prognostic molecular biomarkers would be of use. These biomarkers should detect molecular signatures predicting treatment resistance or the existence of a high risk blast crisis primed cell population (Giustacchini et al., 2017).

Previous studies have investigated gene expression between TKI responders and non-responders (McWeeney et al., 2010, Giustacchini et al., 2017). Unsupervised whole transcriptome analysis revealed no clear clusters between TKI responder and non-responder CML cells (Chapter 4) (Giustacchini et al., 2017). However, gene expression differences have been found in BCR-ABL1⁻ cells between the two groups when a single-cell approach has been used (Giustacchini et al., 2017).

Although, single-cell sequencing is a very powerful technique, it is still expensive and challenging, making it unsuitable for its application in the clinic until more affordable and standardised protocols are developed. However, analysing the gene expression profile of a panel of genes in a bulk population is possible. A previous study investigated the gene expression differences between IM responders and non-responders in unselected CML cells finding no differences between the two groups (Crossman et al., 2005). In contrast, the same group was able to build a gene expression classifier using CD34⁺ enriched cells that correctly predicts IM response in a different set of patients, suggesting that it is necessary to investigate the gene expression of a more primitive cell population for predicting TKI response (McWeeney et al., 2010). However, this classifier did not correctly predict response to nilotinib, another TKI (Patel et al., 2018). As discussed previously (Chapter 4), this could be explained by the age difference between the IM responders and non-responders.

The work presented in this thesis shows the potential of the TKI₁₆₁ signature for predicting IM response by analysing gene expression data (Chapter 4). However, the TKI₁₆₁ classifier requires more tuning (e.g. optimisation of the number of components) and be tested in additional datasets in order to avoid centre, population and treatment bias. The analysis of Patel (Patel et al., 2018) and Giustacchini (Giustacchini et al., 2017) datasets would be of great value as response to TKI was assessed for nilotinib and a mix of TKIs (IM, nilotinib, dasatinib and bosutinib) respectively.

Interestingly, *ERG* expression has been found to have an important role in predicting patient outcome in other types of leukaemias. *ERG* is found overexpressed in AML and a higher expression is associated with poorer outcome (Marcucci et al., 2005). Similarly, higher expression of *ERG* predicts for adverse outcome in T-ALL (Baldus et al., 2006) and acute megakaryoblastic leukaemia (AMKL) (Salek-Ardakani et al., 2009). Intragenic deletions of *ERG* have been found in B-ALL and this constitutes a marker of good prognosis (Clappier et al., 2013, Harvey et al., 2010). Interestingly, *ERG* intragenic deletions were associated with a higher expression of *CHST2*, similarly to the findings of this thesis (Chapter 3). It is also of note, that the presence of *ERG* deletions was associated with a different set of patients than those expressing a BCR-ABL1 like signature in B-ALL (Harvey et al., 2010), supporting the TK independent role of this *ERG* downregulation in CML and, hence, the TKI signatures.

With this in mind, the TKI₁₆₁ signature may have the potential to compliment the EUTOS risk score (Hasford et al., 2011) in predicting TKI response. This might be led by the role of *ERG* as a global haematopoiesis regulator (Loughran et al., 2008, Knudsen et al., 2015). However, more work should be done on this before suggesting its use in a clinical setting.

In conclusion, the work presented in this thesis confirms the existence of a BCR-ABL1 TK independent signature in CML LSCs (TKI signature), which existence has long been assumed in the CML community. Furthermore, this signature affects pathways that have already been successfully targeted in CML for the elimination of CML LSCs, such as p53, MYC (Abraham et al., 2016) and β -catenin (Eiring et al., 2015) and includes genes that have been previously shown to be independent of the BCR-ABL1 TK such as *MIR10A* (Agirre et al., 2008) and *ERG* (Harvey et al., 2010). Additionally, it has been shown that targeting CD33, a member of the TKI signature, using low concentrations of GO in CML CD34⁺ cells is effective and have a limited effect in normal cells. The use of the TKI signature has also shown potential for predicting TKI response in CML CD34⁺ cells.

Taken together, the demonstration of the existence of the TKIi signature can open new avenues for CML treatment and may constitute a biomarker for current therapy.

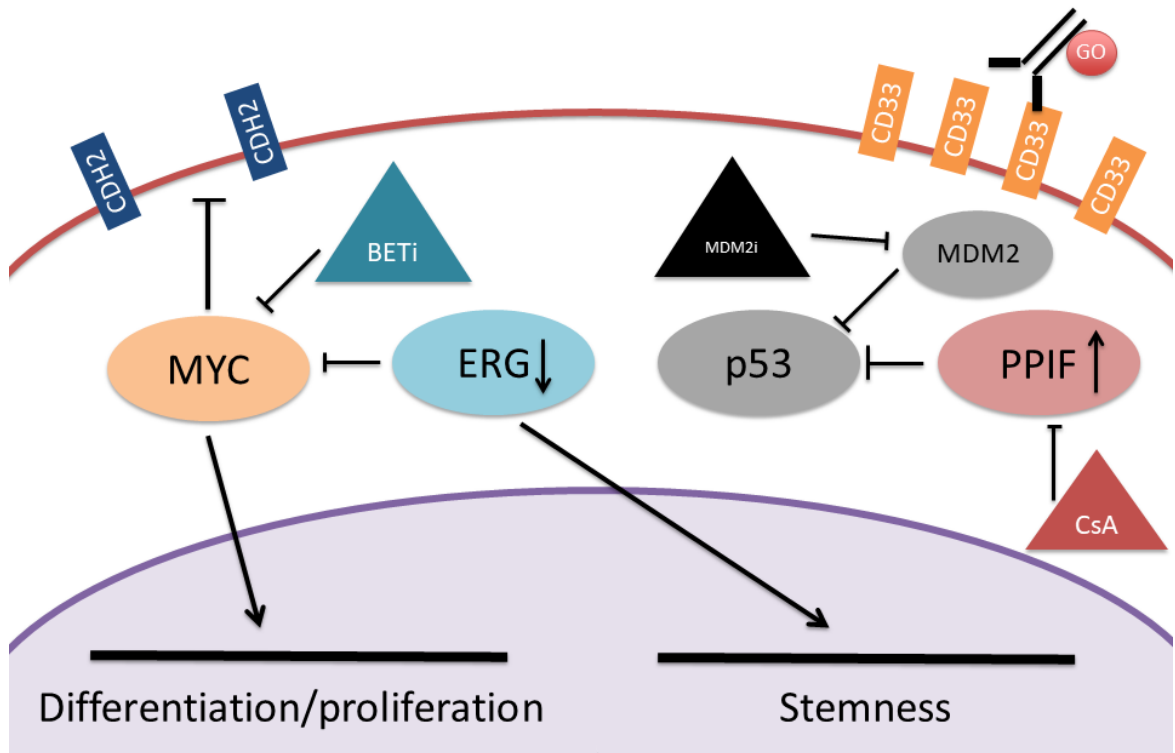


Figure 6-1. Proposed integration of the TKIi signature with the current literature. Soluble proteins are represented with ovals, membrane proteins with rectangles and drugs with triangles. GO is represented as an immunoglobulin. ERG promotes stemness and inhibits the differentiation and proliferation pathways induced by MYC. Therefore, its downregulation activates MYC pathways, which reduce the expression of certain surface proteins, such as N-cadherin (CDH2). MYC can be inhibited using BETi. PPIF, which is upregulated in CML, can inhibit p53. Therefore, its inhibition with cyclosporine A (CsA) may activate p53 pathways in a similar way than MDM2i do. Additionally, surface proteins, like CD33, can be targeted with the use of monoclonal antibodies (like GO).

7 Appendix I

Table 7-1. Primer sequences of the primers used for qPCR gene expression analysis.
Reference (housekeeping) genes are marked in red.

HGNC	Forward primer	Reverse primer	Length
ARHGAP18	GCCACCTCGAGTCGCAGA	GATCGATCAAATGGAGGCTTCTC	89
ASAP2	TCTCCGTGTCGGAATTCGTG	ACGTCCAAAGCCTCCTCGAT	124
ATP5B	TCCATCCTGTCAGGGACTATG	ATCAAACCTGGACGTCCACCAC	110
B2M	TTGTCTTTCAGCAAGGACTGG	ATGCGGCATCTTCAAACCTCC	172
BMP6	TCAACCGCAAGAGCCTTCTG	TCGTAATCCACCAGGTTTAC	130
C10orf10	ACCTCCTCAGCTCCAGGTTG	TCCCGAATTGTGGGCAGATG	91
CCDC159	GTGACTCAGTCTCTGAGCGT	GTCTCCAAGGGCTTCACCTG	125
CD33	ACGTACCTATGTTCCACAGA	AATGGCCCCATGAACCACTC	98
CDH2	AGGCTTCTGGTGAAATCGCA	GCAGTTGTAAACTTCACATTGAG	119
CHST2	ATACTGCTGGCGATGGTAGC	TTCGCTGCTTTCATCCACCT	112
CPNE5	CTGGAAAAGCCCCCTCACGAT	GACACCACCGTTGGACACA	84
DNAH10	GCAATCAACTTTTCACCGCT	TCCCAGCTGTGTACCTAAGGA	136
DUSP18	AAGGAGAGAGCTTTGTTTAAGACTG	ATTGGGGTAGAGGTTAGGCC	147
ENOX2	GAGCTGGAGGGAACCTGATTT	CACTGGCACTACCAAACCTGCA	123
ERG	AGTCGAAAGCTGCTCAACCA	CACTGCCTGGATTTGCAAGG	120
ESAM	CTTCCAGCGTGGTACACCTT	GGACAACACCTGATCCTCCTT	111
GAPDH	ACGGATTTGGTCGTATTGGG	ATTTTGGAGGGATCTGCTC	
GIPC2	ATCAAAAGGTCTGCCACCG	TCAAACATTGTGGTGGCTAAATCA	129
GMPR	AAAGCCTTTGGAGCTGGAGCA	CCGTCCGTTCTCTCAAACA	96
GRB10	CTACCAGGACAAGGTGGAGC	CCTCCTGGTGATTGCGAAGT	98
IGFBP2	CATGGGCGAGGGCACTTGT	TCATCGCCATTGTCTGCAACC	81
MIR10A	ACCCTGTAGATCCGAATTTGTGTAA	AGAGCGGAGTGTTTATGTCAACT	88
MIR21	ATGTTGACTGTTGAATCTCATGGC	TGTCAGACAGCCCATCGAC	50
PLAG1	TGTTAAAGCCCCGCGGTTG	CCCTGCTCCAAACTCTAGCA	114
PPIF	TTACACTGAAGCACGTGGGG	GCCATCCAACCAGTCTGTCT	110
PRPF8	AACGCTCACCACCAAGGAAA	TGCACGTGACTATCCACCAC	105
PTPN7	GCTTCCTGGAGCCTTCTCAG	CAGCCATGAGGTCTGCTGAA	105
RAB38	TGGGATATCGCAGGTCAAGAAA	TCCACTTTGCCACTGCTTCA	124
RASSF9	CCCCCACCCTCAGATCACTT	TGGAGATCTGTTTTATGCCGAGTC	109
RNF180	TCTTTTGGAGGTTTCCGCAG	AACCACCATCCACGGCTATC	114
RNF20	GGTGCTCTTCAACGGAGGAA	TAGTGAGGCATCATCAGTGGC	156
SLC16A10	CATGTGCATTGGCGTCACTT	TGAAGAAAAAGGAGAGTACCTGT	110
TBC1D12	GCGATCACCTGCCCCC	AAAAGTTTCTGGTGAAGAAGTCC	124
TCEB2	GGCTGTACAAGGATGACCAACTCT	ATGCACAGGGCCTCAAAGGT	142
TMEFF1	CAAGAGCATCAACTGCTCAGAATTA	TGGCATGCACATTTCAAACCAT	123
TRIM58	CGGCAGCTACCAGGTAAAGC	GCTGCCTCTGCATTTCCACT	134
TYW1	ATTGTCATCAAGACGCGGGC	GTTGCGAATCCCTTCGCTGTT	167
UBASH3B	ACCATCAAGCATGGATCGGC	CCGACATGGGAGAATAACCAGT	134
VMP1	GGGTTCCGGTTGTCTGGAG	CAGTAACTCTTGAGGAGCCGC	112
ZDBF2	ATAAGAAGGGAGAGCGCCCG	GAATACTCAAGCTGGAGCAGAAAA	140
ZMAT3	GATGCCTCCTTCAGTTCCCC	GCTCTGAGGATTCCGAGAATGA	115

8 Appendix II

Tables containing all the over and under-represented pathways in each of the contrasts of the RNAseq experiment described in section 5.2.28. Calculations were performed by CPDB (Kamburov et al., 2013).

Table 8-1. GO vs NDC.

Pathway	Source	q-value
Hemostasis	Reactome	4.8E-12
Platelet activation, signaling and aggregation	Reactome	1.2E-09
p53 signaling pathway - Homo sapiens (human)	KEGG	3.0E-09
Platelet degranulation	Reactome	3.3E-09
Response to elevated platelet cytosolic Ca ²⁺	Reactome	4.8E-09
Formation of Fibrin Clot (Clotting Cascade)	Reactome	1.9E-08
Cell surface interactions at the vascular wall	Reactome	9.0E-07
Cytokine-cytokine receptor interaction - Homo sapiens (human)	KEGG	2.6E-06
Common Pathway of Fibrin Clot Formation	Reactome	4.0E-06
Complement and coagulation cascades - Homo sapiens (human)	KEGG	7.5E-06
Malaria - Homo sapiens (human)	KEGG	1.3E-04
Hematopoietic cell lineage - Homo sapiens (human)	KEGG	1.7E-04
Platelet Adhesion to exposed collagen	Reactome	8.3E-04
TP53 Regulates Transcription of Cell Death Genes	Reactome	9.9E-04
hemoglobins chaperone	BioCarta	1.0E-03
Peptide ligand-binding receptors	Reactome	3.6E-03
TNFs bind their physiological receptors	Reactome	3.6E-03
Erythrocytes take up oxygen and release carbon dioxide	Reactome	3.9E-03
Melanoma - Homo sapiens (human)	KEGG	4.8E-03
GP1b-IX-V activation signalling	Reactome	5.2E-03
Signal Transduction	Reactome	6.0E-03
Platelet activation - Homo sapiens (human)	KEGG	6.0E-03
TP53 Regulates Transcription of Death Receptors and Ligands	Reactome	6.0E-03
Scavenging of heme from plasma	Reactome	6.0E-03
Class A/1 (Rhodopsin-like receptors)	Reactome	6.0E-03
African trypanosomiasis - Homo sapiens (human)	KEGG	6.3E-03
Erythrocytes take up carbon dioxide and release oxygen	Reactome	7.0E-03
O ₂ /CO ₂ exchange in erythrocytes	Reactome	7.0E-03
ECM-receptor interaction - Homo sapiens (human)	KEGG	8.7E-03
Amino acid synthesis and interconversion (transamination)	Reactome	1.0E-02
p53 signaling pathway	BioCarta	1.1E-02
PI3K/AKT Signaling in Cancer	Reactome	1.2E-02
Intrinsic Pathway of Fibrin Clot Formation	Reactome	1.2E-02
GPCR ligand binding	Reactome	1.2E-02
Signaling by GPCR	Reactome	1.2E-02
Chemokine receptors bind chemokines	Reactome	1.4E-02

Apoptosis - Homo sapiens (human)	KEGG	1.5E-02
Rap1 signaling pathway - Homo sapiens (human)	KEGG	1.7E-02
Transcriptional Regulation by TP53	Reactome	2.3E-02
Thyroid cancer - Homo sapiens (human)	KEGG	2.3E-02
Glioma - Homo sapiens (human)	KEGG	2.5E-02
MAPK family signaling cascades	Reactome	2.5E-02
Extracellular matrix organization	Reactome	2.5E-02
regulation of bad phosphorylation	BioCarta	2.5E-02
p53-Dependent G1 DNA Damage Response	Reactome	2.5E-02
p53-Dependent G1/S DNA damage checkpoint	Reactome	2.5E-02
GPCR downstream signalling	Reactome	2.5E-02
TNFR2 non-canonical NF-kB pathway	Reactome	2.5E-02
Bladder cancer - Homo sapiens (human)	KEGG	2.5E-02
hypoxia and p53 in the cardiovascular system	BioCarta	2.8E-02
Immune System	Reactome	2.8E-02
G1/S DNA Damage Checkpoints	Reactome	3.0E-02
Basal cell carcinoma - Homo sapiens (human)	KEGG	3.2E-02
PI3K-Akt signaling pathway - Homo sapiens (human)	KEGG	3.2E-02
Neutrophil degranulation	Reactome	3.4E-02
Constitutive Signaling by Aberrant PI3K in Cancer	Reactome	3.6E-02

Table 8-2. IM vs NDC.

Pathway	Source	q-value
Complement and coagulation cascades - Homo sapiens (human)	KEGG	0.010
GPCR ligand binding	Reactome	0.010
Formation of Fibrin Clot (Clotting Cascade)	Reactome	0.010
Signaling by GPCR	Reactome	0.021
Class A/1 (Rhodopsin-like receptors)	Reactome	0.037

Table 8-3. Combination treatment (GO+IM) vs NDC.

Pathway	Source	q-value
Hemostasis	Reactome	9.4E-07
Response to elevated platelet cytosolic Ca ²⁺	Reactome	7.0E-06
Platelet degranulation	Reactome	6.0E-05
Cytokine-cytokine receptor interaction - Homo sapiens	KEGG	1.6E-04
Platelet activation, signaling and aggregation	Reactome	1.6E-04
Chemokine receptors bind chemokines	Reactome	1.0E-03
Nucleotide-like (purinergic) receptors	Reactome	4.2E-03
G alpha (i) signalling events	Reactome	1.1E-02
Complement and coagulation cascades - Homo sapiens	KEGG	1.3E-02
RHO GTPases activate PKNs	Reactome	1.3E-02
GPCR ligand binding	Reactome	1.3E-02

Smooth Muscle Contraction	Reactome	1.3E-02
Class A/1 (Rhodopsin-like receptors)	Reactome	1.3E-02
p53 signaling pathway - Homo sapiens (human)	KEGG	1.3E-02
Systemic lupus erythematosus - Homo sapiens (human)	KEGG	1.7E-02
P2Y receptors	Reactome	1.7E-02
Neutrophil degranulation	Reactome	2.3E-02
RAB geranylgeranylation	Reactome	2.3E-02
cdk regulation of dna replication	BioCarta	2.7E-02
Reproduction	Reactome	2.9E-02
DNA methylation	Reactome	2.9E-02
Meiotic recombination	Reactome	3.7E-02
Amyloid fiber formation	Reactome	3.7E-02
Act PKN1 stimulates transcription of AR reg genes KLK2 and KLK3	Reactome	3.7E-02
Syndecan interactions	Reactome	4.6E-02
Signal Transduction	Reactome	4.6E-02
G2/M Checkpoints	Reactome	4.6E-02
Synthesis of 5-eicosatetraenoic acids	Reactome	4.7E-02
Immunoreg interactions between Lymphoid and non-Lymphoid cell	Reactome	8.4E-02
Acyl chain remodelling of PE	Reactome	8.4E-02
RNA Polymerase I Promoter Opening	Reactome	8.4E-02
Cell Cycle, Mitotic	Reactome	8.4E-02
e2f1 destruction pathway	BioCarta	8.5E-02
RHO GTPase Effectors	Reactome	8.5E-02
TP53 Reg Transcription of Genes Involved in G1 Cell Cycle Arrest	Reactome	8.5E-02
Arachidonic acid metabolism - Homo sapiens	KEGG	8.5E-02
Mitotic G1-G1/S phases	Reactome	8.5E-02
Megakaryocyte development and platelet production	Reactome	8.9E-02
Cell Cycle	Reactome	8.9E-02
B-WICH complex positively regulates rRNA expression	Reactome	8.9E-02
Meiosis	Reactome	9.6E-02
Condensation of Prophase Chromosomes	Reactome	9.6E-02
Mitotic Prophase	Reactome	9.6E-02
Synthesis of 12-eicosatetraenoic acid derivatives	Reactome	9.6E-02
Synthesis of 15-eicosatetraenoic acid derivatives	Reactome	9.6E-02
eicosanoid metabolism	BioCarta	1.1E-01
PRC2 methylates histones and DNA	Reactome	1.2E-01
SIRT1 negatively regulates rRNA expression	Reactome	1.2E-01
TNF receptor superfamily mediating non-canonical NF-kB pathway	Reactome	1.4E-01
Peptide ligand-binding receptors	Reactome	1.4E-01
RUNX1 reg megakaryocyte differentiation	Reactome	1.4E-01

and platelet function		
Cell surface interactions at the vascular wall	Reactome	1.4E-01
Zinc transporters	Reactome	1.4E-01
Innate Immune System	Reactome	1.4E-01
rb tumor suppressor/checkpoint signaling in response to dna damage	BioCarta	1.4E-01
RHO GTPases activate IQGAPs	Reactome	1.4E-01
TNFR2 non-canonical NF-kB pathway	Reactome	1.5E-01
Formation of the beta-catenin: TCF transactivating complex	Reactome	1.5E-01
Cell cycle - Homo sapiens (human)	KEGG	1.5E-01
Transport of small molecules	Reactome	1.5E-01
DNA Damage/Telomere Stress Induced Senescence	Reactome	1.5E-01
G1/S Transition	Reactome	1.5E-01
Packaging Of Telomere Ends	Reactome	1.5E-01
Purine metabolism - Homo sapiens (human)	KEGG	1.5E-01
Senescence-Associated Secretory Phenotype (SASP)	Reactome	1.6E-01
Transcriptional misregulation in cancer - Homo sapiens	KEGG	1.6E-01
RNA Polymerase I Chain Elongation	Reactome	1.6E-01
RHO GTPases activate CIT	Reactome	1.6E-01
cell cycle: g2/m checkpoint	BioCarta	1.6E-01
cell cycle: g1/s check point	BioCarta	1.6E-01
Telomere Maintenance	Reactome	1.7E-01
ERCC6 (CSB) and EHMT2 (G9a) positively regulate rRNA expression	Reactome	1.7E-01
Transcriptional regulation by small RNAs	Reactome	1.8E-01
Phagosome - Homo sapiens (human)	KEGG	1.8E-01
Biosynthesis of DPAn-3 SPMs	Reactome	1.8E-01
Biosynthesis of DPA-derived SPMs	Reactome	1.8E-01
Disinhibition of SNARE formation	Reactome	1.8E-01
Immune System	Reactome	1.8E-01
Cholesterol metabolism - Homo sapiens (human)	KEGG	1.8E-01
Activation of the pre-replicative complex	Reactome	1.8E-01

Table 8-4. Combination treatment (GO+IM) vs IM.

Pathway	Source	q-value
Cytokine-cytokine receptor interaction - Homo sapiens (human)	KEGG	3.2E-07
Response to elevated platelet cytosolic Ca ²⁺	Reactome	5.7E-07
Hemostasis	Reactome	1.6E-06
Platelet degranulation	Reactome	1.9E-06
Chemokine receptors bind chemokines	Reactome	1.4E-05
Platelet activation, signaling and aggregation	Reactome	1.8E-05
p53 signaling pathway - Homo sapiens (human)	KEGG	1.2E-04
Neutrophil degranulation	Reactome	1.6E-04

GPCR ligand binding	Reactome	1.4E-03
G alpha (i) signalling events	Reactome	1.4E-03
Class A/1 (Rhodopsin-like receptors)	Reactome	1.5E-03
Chemokine signaling pathway - Homo sapiens (human)	KEGG	3.6E-03
RHO GTPases activate PKNs	Reactome	3.6E-03
Smooth Muscle Contraction	Reactome	7.3E-03
Cell cycle - Homo sapiens (human)	KEGG	7.3E-03
Innate Immune System	Reactome	8.4E-03
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	Reactome	1.1E-02
Synthesis of 5-eicosatetraenoic acids	Reactome	1.6E-02
Reproduction	Reactome	1.7E-02
TP53 Regulates Transcription of Genes Involved in G1 Cell Cycle Arrest	Reactome	1.8E-02
Nucleotide-like (purinergic) receptors	Reactome	2.0E-02
P2Y receptors	Reactome	2.0E-02
e2f1 destruction pathway	BioCarta	2.0E-02
Mitotic G1-G1/S phases	Reactome	2.1E-02
Signal Transduction	Reactome	2.3E-02
Cell surface interactions at the vascular wall	Reactome	3.3E-02
cell cycle: g1/s check point	BioCarta	3.3E-02
Immune System	Reactome	3.4E-02
Malaria - Homo sapiens (human)	KEGG	3.4E-02
Cell Cycle	Reactome	3.4E-02
Peptide ligand-binding receptors	Reactome	3.4E-02
p53 signaling pathway	BioCarta	3.4E-02
Complement and coagulation cascades - Homo sapiens	KEGG	3.5E-02
cyclins and cell cycle regulation	BioCarta	3.5E-02
cdk regulation of dna replication	BioCarta	3.5E-02
Cell Cycle, Mitotic	Reactome	3.5E-02
G1/S Transition	Reactome	3.5E-02
Erythrocytes take up carbon dioxide and release oxygen	Reactome	3.5E-02
O2/CO2 exchange in erythrocytes	Reactome	3.5E-02
Transcriptional misregulation in cancer - Homo sapiens	KEGG	3.5E-02
Activated PKN1 stimulates transcription of AR regulated genes KLK2 and KLK3	Reactome	3.9E-02
Lysosphingolipid and LPA receptors	Reactome	3.9E-02
Systemic lupus erythematosus - Homo sapiens (human)	KEGG	4.4E-02
Meiosis	Reactome	4.6E-02
Transcriptional regulation by small RNAs	Reactome	4.8E-02
Meiotic recombination	Reactome	4.9E-02
RHO GTPases activate PAKs	Reactome	5.2E-02
Viral carcinogenesis - Homo sapiens (human)	KEGG	5.4E-02
Alcoholism - Homo sapiens (human)	KEGG	5.8E-02

DNA methylation	Reactome	5.9E-02
TNFR2 non-canonical NF-kB pathway	Reactome	6.1E-02
RUNX1 regulates megakaryocyte differentiation and platelet function genes	Reactome	6.3E-02
Activation of E2F1 target genes at G1/S	Reactome	6.3E-02
G1/S-Specific Transcription	Reactome	6.3E-02
Transferrin endocytosis and recycling	Reactome	6.3E-02
Mitotic Prophase	Reactome	6.3E-02
RAB geranylgeranylation	Reactome	6.6E-02
Scavenging of heme from plasma	Reactome	7.2E-02
PI3K-Akt signaling pathway - Homo sapiens (human)	KEGG	7.3E-02
Amyloid fiber formation	Reactome	7.4E-02
Phagosome - Homo sapiens (human)	KEGG	7.4E-02
Condensation of Prophase Chromosomes	Reactome	7.5E-02
TNF receptor superfamily (TNFSF) members mediating non-canonical NF-kB pathway	Reactome	7.5E-02
Cholesterol biosynthesis	Reactome	7.7E-02
Hematopoietic cell lineage - Homo sapiens (human)	KEGG	7.8E-02
RNA Polymerase I Promoter Opening	Reactome	8.1E-02
Formation of Fibrin Clot (Clotting Cascade)	Reactome	8.1E-02
RHO GTPases activate CIT	Reactome	8.6E-02
cell cycle: g2/m checkpoint	BioCarta	8.6E-02
Adrenaline,noradrenaline inhibits insulin secretion	Reactome	8.6E-02
regulation of p27 phosphorylation during cell cycle progression	BioCarta	9.3E-02
Serotonergic synapse - Homo sapiens (human)	KEGG	9.4E-02
Signaling by GPCR	Reactome	9.9E-02
Prostate cancer - Homo sapiens (human)	KEGG	1.1E-01
G2/M Checkpoints	Reactome	1.1E-01
Metabolism of water-soluble vitamins and cofactors	Reactome	1.1E-01
Ferroptosis - Homo sapiens (human)	KEGG	1.1E-01
S Phase	Reactome	1.1E-01
Senescence-Associated Secretory Phenotype (SASP)	Reactome	1.1E-01
Transcription of E2F targets under negative control by p107 (RBL1) and p130 (RBL2) in complex with HDAC1	Reactome	1.1E-01
rb tumor suppressor/checkpoint signaling in response to dna damage	BioCarta	1.1E-01
Beta-catenin independent WNT signaling	Reactome	1.2E-01
eicosanoid metabolism	BioCarta	1.2E-01
G1/S DNA Damage Checkpoints	Reactome	1.3E-01
G0 and Early G1	Reactome	1.4E-01
ROS, RNS production in phagocytes	Reactome	1.4E-01
Arginine and proline metabolism - Homo sapiens (human)	KEGG	1.4E-01
influence of ras and rho proteins on g1 to s transition	BioCarta	1.4E-01
cyclin e destruction pathway	BioCarta	1.4E-01
Platelet activation - Homo sapiens (human)	KEGG	1.4E-01

GPCR downstream signalling	Reactome	1.5E-01
Ephrin signaling	Reactome	1.5E-01
ADP signalling through P2Y purinoceptor 12	Reactome	1.5E-01

9 Appendix III

Tables containing all the over and under-represented GO-terms in each of the contrasts of the RNAseq experiment described in section 5.2.28. Calculations were performed by CPDB (Kamburov et al., 2013).

Table 9-1. GO vs NDC.

Name	ID	q-value
blood coagulation	GO:0007596	9.27E-11
response to wounding	GO:0009611	2.41E-09
regulation of body fluid levels	GO:0050878	5.29E-09
platelet activation	GO:0030168	5.29E-09
regulation of signaling	GO:0023051	6.75E-07
regulation of response to stimulus	GO:0048583	6.75E-07
cell surface receptor signaling pathway	GO:0007166	7.31E-07
positive regulation of biological process	GO:0048518	7.31E-07
regulation of hemostasis	GO:1900046	8.50E-07
exocytosis	GO:0006887	1.57E-06
regulation of multicellular organismal process	GO:0051239	1.97E-06
vesicle lumen	GO:0031983	1.23E-05
intrinsic component of plasma membrane	GO:0031226	1.23E-05
apoptotic signaling pathway	GO:0097190	1.38E-05
extracellular vesicle	GO:1903561	1.88E-05
vesicle	GO:0031982	1.88E-05
intracellular signal transduction	GO:0035556	3.13E-05
anchored component of membrane	GO:0031225	1.88E-05
myeloid leukocyte migration	GO:0097529	4.99E-05
regulation of signaling receptor activity	GO:0010469	4.99E-05
response to external biotic stimulus	GO:0043207	6.04E-05
cellular response to chemical stimulus	GO:0070887	7.51E-05
haptoglobin binding	GO:0031720	0.0002339
signaling receptor binding	GO:0005102	0.0002339
leukocyte chemotaxis	GO:0030595	0.00013634
response to bacterium	GO:0009617	0.00023817
regulation of immune system process	GO:0002682	0.0002661
regulation of localization	GO:0032879	0.00029603
cell-cell adhesion	GO:0098609	0.0005863
oxidoreductase activity, acting on peroxide as acceptor	GO:0016684	0.0012507
chemotaxis	GO:0006935	0.00080582
response to oxygen-containing compound	GO:1901700	0.00096759
blood coagulation, fibrin clot formation	GO:0072378	0.0009968
reactive oxygen species metabolic process	GO:0072593	0.0009968
nephron development	GO:0072006	0.00107965
circulatory system process	GO:0003013	0.00108591

mononuclear cell proliferation	GO:0032943	0.00108591
response to organic substance	GO:0010033	0.00115475
collagen binding	GO:0005518	0.00268898
regulation of cellular process	GO:0050794	0.00127104
regulation of cell adhesion	GO:0030155	0.00139654
negative regulation of biological process	GO:0048519	0.00184145
cell differentiation	GO:0030154	0.00190798
glycosaminoglycan binding	GO:0005539	0.00390458
programmed cell death	GO:0012501	0.00251474
regulation of developmental process	GO:0050793	0.00255447
transport	GO:0006810	0.00321951
oxygen binding	GO:0019825	0.00697998
muscle system process	GO:0003012	0.00429025
defense response	GO:0006952	0.00496074
blood coagulation, intrinsic pathway	GO:0007597	0.00496074
glomerulus development	GO:0032835	0.00602575
heparin binding	GO:0008201	0.01019035
G-protein coupled receptor signaling pathway	GO:0007186	0.00939751
cellular oxidant detoxification	GO:0098869	0.00940964
cytoplasmic vesicle part	GO:0044433	0.0141549
cell junction organization	GO:0034330	0.01059913
TRAIL binding	GO:0045569	0.01633244
tissue development	GO:0009888	0.01108119
extracellular structure organization	GO:0043062	0.01110228
transferase activity, transferring nitrogenous groups	GO:0016769	0.01633244
external side of plasma membrane	GO:0009897	0.0141549
intracellular vesicle	GO:0097708	0.0141549
membrane microdomain	GO:0098857	0.0141549
blood vessel development	GO:0001568	0.01213232
humoral immune response	GO:0006959	0.0149378
hemoglobin binding	GO:0030492	0.02227085
cell aging	GO:0007569	0.01680404
signal transduction by protein phosphorylation	GO:0023014	0.01808314
response to toxic substance	GO:0009636	0.01943658
lipopolysaccharide binding	GO:0001530	0.02711697
integral component of membrane	GO:0016021	0.03011549
response to drug	GO:0042493	0.02735792
cofactor metabolic process	GO:0051186	0.02735792
cell activation involved in immune response	GO:0002263	0.02829326
cytokine production involved in immune response	GO:0002367	0.02829326
ectopic germ cell programmed cell death	GO:0035234	0.02898383
cell adhesion molecule binding	GO:0050839	0.0412482
smooth muscle cell proliferation	GO:0048659	0.03217981
leukocyte degranulation	GO:0043299	0.03291973
signaling receptor activity	GO:0038023	0.04140678

renal system process	GO:0003014	0.03332496
cell development	GO:0048468	0.03332496
cellular chemical homeostasis	GO:0055082	0.03332496
heme binding	GO:0020037	0.04140678
regulation of catalytic activity	GO:0050790	0.03510374
scaffold protein binding	GO:0097110	0.04140678
dendritic cell migration	GO:0036336	0.03708241
cell-cell adherens junction	GO:0005913	0.04834098
response to transforming growth factor beta	GO:0071559	0.03798788
system development	GO:0048731	0.03833257
mononuclear cell migration	GO:0071674	0.04148662
response to xenobiotic stimulus	GO:0009410	0.04406183
calmodulin binding	GO:0005516	0.05334254

Table 9-2. IM vs NDC.

Name	ID	q-value
regulation of multicellular organismal process	GO:0051239	3.3E-06
regulation of response to stimulus	GO:0048583	5.6E-05
cell surface receptor signaling pathway	GO:0007166	5.6E-05
response to organic substance	GO:0010033	5.3E-04
negative regulation of biological process	GO:0048519	5.5E-04
regulation of hemostasis	GO:1900046	5.7E-04
regulation of signaling	GO:0023051	1.4E-03
blood coagulation, fibrin clot formation	GO:0072378	1.9E-03
regulation of developmental process	GO:0050793	2.3E-03
positive regulation of biological process	GO:0048518	3.2E-03
cellular response to chemical stimulus	GO:0070887	8.3E-03
leading edge membrane	GO:0031256	2.7E-02
chemotaxis	GO:0006935	9.7E-03
trabecula morphogenesis	GO:0061383	9.7E-03
tissue development	GO:0009888	1.8E-02
system development	GO:0048731	1.8E-02
response to nerve growth factor	GO:1990089	1.9E-02
regulation of localization	GO:0032879	1.9E-02
isoprenoid binding	GO:0019840	6.1E-02
regulation of cellular process	GO:0050794	1.9E-02
cardiac muscle tissue growth	GO:0055017	2.2E-02
signal transduction by protein phosphorylation	GO:0023014	2.2E-02
heart growth	GO:0060419	2.6E-02
animal organ development	GO:0048513	2.6E-02
secondary metabolite biosynthetic process	GO:0044550	2.6E-02
regulation of synapse structure or activity	GO:0050803	2.6E-02
G-protein coupled receptor signaling pathway	GO:0007186	2.6E-02
G-protein alpha-subunit binding	GO:0001965	6.5E-02

animal organ morphogenesis	GO:0009887	2.8E-02
nuclear membrane	GO:0031965	5.5E-02
cell differentiation	GO:0030154	2.8E-02
regulation of locomotion	GO:0040012	3.0E-02
intrinsic component of plasma membrane	GO:0031226	5.5E-02
blood coagulation	GO:0007596	3.0E-02
regulation of signaling receptor activity	GO:0010469	3.0E-02
intracellular signal transduction	GO:0035556	3.0E-02
proteinaceous extracellular matrix	GO:0005578	5.5E-02
blood vessel development	GO:0001568	3.2E-02
tube morphogenesis	GO:0035239	3.3E-02
negative regulation of molecular function	GO:0044092	3.4E-02
response to oxygen-containing compound	GO:1901700	3.4E-02
tube development	GO:0035295	3.4E-02
response to wounding	GO:0009611	3.4E-02
regulation of binding	GO:0051098	3.5E-02
regulation of cellular component biogenesis	GO:0044087	3.7E-02
vesicle	GO:0031982	7.2E-02
developmental growth involved in morphogenesis	GO:0060560	4.6E-02
cellular response to endogenous stimulus	GO:0071495	4.6E-02

Table 9-3. Combination (GO+IM) vs NDC.

Name	ID	q-value
cellular response to chemical stimulus	GO:0070887	1.5E-05
regulation of immune system process	GO:0002682	1.5E-05
response to wounding	GO:0009611	1.7E-05
response to oxygen-containing compound	GO:1901700	1.7E-05
response to organic substance	GO:0010033	1.7E-05
negative regulation of biological process	GO:0048519	1.7E-05
extracellular vesicle	GO:1903561	3.8E-05
regulation of localization	GO:0032879	1.8E-05
regulation of response to stimulus	GO:0048583	2.2E-05
regulation of multicellular organismal process	GO:0051239	4.1E-05
exocytosis	GO:0006887	5.0E-05
intrinsic component of plasma membrane	GO:0031226	1.3E-04
cell activation involved in immune response	GO:0002263	6.3E-05
defense response	GO:0006952	7.9E-05
cellular chemical homeostasis	GO:0055082	9.6E-05
integral component of membrane	GO:0016021	2.0E-04
blood coagulation	GO:0007596	1.7E-04
regulation of signaling	GO:0023051	2.0E-04
leukocyte mediated immunity	GO:0002443	2.6E-04
vesicle	GO:0031982	3.9E-04

cytoplasmic vesicle part	GO:0044433	7.0E-04
chemotaxis	GO:0006935	9.0E-04
response to bacterium	GO:0009617	9.0E-04
leukocyte degranulation	GO:0043299	9.5E-04
regulation of locomotion	GO:0040012	1.2E-03
leukocyte chemotaxis	GO:0030595	1.2E-03
oxidoreductase activity, acting on single donors with incorporation of O2	GO:0016701	1.0E-02
response to external biotic stimulus	GO:0043207	1.3E-03
mitotic cell cycle process	GO:1903047	1.3E-03
negative regulation of molecular function	GO:0044092	1.5E-03
cell surface receptor signaling pathway	GO:0007166	1.5E-03
homeostatic process	GO:0042592	1.6E-03
intracellular signal transduction	GO:0035556	2.1E-03
regulation of hemostasis	GO:1900046	2.1E-03
platelet activation	GO:0030168	2.2E-03
positive regulation of biological process	GO:0048518	2.9E-03
secretory granule membrane	GO:0030667	6.0E-03
cell differentiation	GO:0030154	3.3E-03
vesicle lumen	GO:0031983	6.0E-03
regulation of developmental process	GO:0050793	3.8E-03
regulation of body fluid levels	GO:0050878	3.9E-03
protein dimerization activity	GO:0046983	2.8E-02
regulation of binding	GO:0051098	4.2E-03
humoral immune response	GO:0006959	5.2E-03
system development	GO:0048731	6.2E-03
external side of plasma membrane	GO:0009897	1.0E-02
circulatory system process	GO:0003013	6.6E-03
regulation of anatomical structure size	GO:0090066	6.7E-03
transport	GO:0006810	7.3E-03
mitotic cell cycle	GO:0000278	7.3E-03
animal organ development	GO:0048513	7.3E-03
myeloid leukocyte migration	GO:0097529	7.3E-03
cellular oxidant detoxification	GO:0098869	7.3E-03
cell adhesion molecule binding	GO:0050839	5.3E-02
nucleosome	GO:0000786	2.2E-02
response to hormone	GO:0009725	1.3E-02
response to drug	GO:0042493	1.5E-02
somatic diversification immune receptors via germline recomb single locus	GO:0002562	1.8E-02
intracellular vesicle	GO:0097708	3.0E-02
identical protein binding	GO:0042802	8.1E-02
immunoglobulin production	GO:0002377	1.8E-02
protein secretion	GO:0009306	1.9E-02
response to inorganic substance	GO:0010035	1.9E-02

hydrolase activity, acting on acid anhydrides	GO:0016817	8.1E-02
organic acid metabolic process	GO:0006082	2.0E-02
tube morphogenesis	GO:0035239	2.1E-02
tissue development	GO:0009888	2.4E-02
cell cycle phase transition	GO:0044770	2.6E-02
mononuclear cell migration	GO:0071674	2.6E-02
somatic diversification of immunoglobulins	GO:0016445	2.6E-02
G-protein coupled receptor signaling pathway	GO:0007186	2.7E-02
innate immune response	GO:0045087	2.8E-02
fibroblast growth factor production	GO:0090269	2.9E-02
regulation of fibroblast growth factor production	GO:0090270	2.9E-02
regulation of cellular process	GO:0050794	2.9E-02
regulation of hormone levels	GO:0010817	3.0E-02
multicellular organism aging	GO:0010259	3.0E-02
blood coagulation, fibrin clot formation	GO:0072378	3.9E-02
sphingolipid binding	GO:0046625	1.7E-01
tube development	GO:0035295	4.1E-02
neuron cellular homeostasis	GO:0070050	4.1E-02
adaptive immune response	GO:0002250	4.1E-02
cellular response to endogenous stimulus	GO:0071495	4.1E-02
reactive oxygen species metabolic process	GO:0072593	4.4E-02
ligase activity, forming carbon-nitrogen bonds	GO:0016879	1.8E-01
regulation of signaling receptor activity	GO:0010469	4.5E-02
response to radiation	GO:0009314	4.5E-02
response to acid chemical	GO:0001101	4.5E-02
mating	GO:0007618	4.7E-02
complement activation	GO:0006956	4.7E-02
small molecule catabolic process	GO:0044282	4.9E-02
vesicle membrane	GO:0012506	1.2E-01
apoptotic signaling pathway	GO:0097190	5.0E-02
response to extracellular stimulus	GO:0009991	5.0E-02
proteoglycan binding	GO:0043394	1.8E-01
enzyme inhibitor activity	GO:0004857	1.8E-01
cell-cell adherens junction	GO:0005913	1.3E-01
signaling receptor binding	GO:0005102	1.8E-01
angiogenesis	GO:0001525	6.2E-02
sulfur compound metabolic process	GO:0006790	6.4E-02
cell-cell adhesion	GO:0098609	6.9E-02
protein to membrane docking	GO:0022615	7.2E-02
oxidoreductase activity, acting on superoxide radicals as acceptor	GO:0016721	2.0E-01
elastic fiber	GO:0071953	1.5E-01
muscle thin filament tropomyosin	GO:0005862	1.5E-01
nucleoside binding	GO:0001882	2.0E-01
lipid localization	GO:0010876	7.3E-02

epithelium migration	GO:0090132	7.6E-02
anion binding	GO:0043168	2.1E-01
type B pancreatic cell proliferation	GO:0044342	7.9E-02
keratinocyte proliferation	GO:0043616	8.4E-02
smooth muscle cell proliferation	GO:0048659	8.5E-02
mating behavior	GO:0007617	8.5E-02
carbon-sulfur lyase activity	GO:0016846	2.2E-01

Table 9-4. Combination treatment (GO+IM) vs IM.

Name	ID	q-value
extracellular vesicle	GO:1903561	1.4E-08
intrinsic component of plasma membrane	GO:0031226	6.8E-08
integral component of membrane	GO:0016021	7.9E-07
exocytosis	GO:0006887	8.1E-06
cell activation involved in immune response	GO:0002263	8.1E-06
leukocyte mediated immunity	GO:0002443	1.1E-05
regulation of response to stimulus	GO:0048583	1.8E-05
regulation of immune system process	GO:0002682	1.8E-05
leukocyte degranulation	GO:0043299	3.2E-05
chemotaxis	GO:0006935	3.7E-05
vesicle	GO:0031982	2.9E-05
cell surface receptor signaling pathway	GO:0007166	4.2E-05
cellular response to chemical stimulus	GO:0070887	4.2E-05
leukocyte chemotaxis	GO:0030595	4.2E-05
response to wounding	GO:0009611	5.7E-05
regulation of localization	GO:0032879	1.1E-04
vesicle lumen	GO:0031983	9.3E-05
regulation of signaling receptor activity	GO:0010469	1.2E-04
defense response	GO:0006952	1.6E-04
response to oxygen-containing compound	GO:1901700	2.2E-04
external side of plasma membrane	GO:0009897	2.1E-04
myeloid leukocyte migration	GO:0097529	2.5E-04
regulation of signaling	GO:0023051	3.3E-04
signaling receptor activity	GO:0038023	2.6E-03
response to organic substance	GO:0010033	6.1E-04
regulation of multicellular organismal process	GO:0051239	6.4E-04
cellular chemical homeostasis	GO:0055082	6.6E-04
intracellular signal transduction	GO:0035556	6.9E-04
protein dimerization activity	GO:0046983	4.2E-03
mitotic cell cycle process	GO:1903047	1.0E-03
cytoplasmic vesicle part	GO:0044433	1.4E-03
platelet activation	GO:0030168	1.5E-03
response to bacterium	GO:0009617	1.6E-03
blood coagulation	GO:0007596	1.6E-03

regulation of locomotion	GO:0040012	1.9E-03
cell differentiation	GO:0030154	2.5E-03
mononuclear cell migration	GO:0071674	2.8E-03
response to external biotic stimulus	GO:0043207	3.0E-03
regulation of developmental process	GO:0050793	3.0E-03
secretory granule membrane	GO:0030667	4.1E-03
extracellular structure organization	GO:0043062	6.2E-03
cellular oxidant detoxification	GO:0098869	6.3E-03
cell cycle phase transition	GO:0044770	6.3E-03
identical protein binding	GO:0042802	3.2E-02
somatic diversification of immunoglobulins	GO:0016445	7.1E-03
negative regulation of biological process	GO:0048519	7.4E-03
system development	GO:0048731	7.4E-03
immunoglobulin production	GO:0002377	7.6E-03
proteinaceous extracellular matrix	GO:0005578	1.1E-02
humoral immune response	GO:0006959	8.2E-03
somatic diversification of immune receptors via germline recomb single locus	GO:0002562	8.7E-03
mitotic cell cycle	GO:0000278	8.7E-03
heparin binding	GO:0008201	3.8E-02
regulation of anatomical structure size	GO:0090066	1.1E-02
oxidoreductase activity, acting on single donors with incorporation of O ₂	GO:0016701	3.8E-02
positive regulation of biological process	GO:0048518	1.1E-02
apoptotic signaling pathway	GO:0097190	1.1E-02
cell-cell adhesion	GO:0098609	1.1E-02
homeostatic process	GO:0042592	1.2E-02
lymphocyte migration	GO:0072676	1.3E-02
intracellular vesicle	GO:0097708	2.5E-02
glycosaminoglycan binding	GO:0005539	5.1E-02
tissue development	GO:0009888	1.8E-02
anchored component of membrane	GO:0031225	2.7E-02
collagen binding	GO:0005518	5.3E-02
signaling receptor binding	GO:0005102	5.3E-02
animal organ development	GO:0048513	2.2E-02
regulation of cellular component size	GO:0032535	2.3E-02
myofilament	GO:0036379	3.5E-02
regulation of body fluid levels	GO:0050878	2.8E-02
oxidoreductase activity, acting on peroxide as acceptor	GO:0016684	6.5E-02
G-protein coupled receptor signaling pathway	GO:0007186	3.0E-02
adaptive immune response	GO:0002250	3.2E-02
regulation of binding	GO:0051098	3.2E-02
protein secretion	GO:0009306	3.6E-02
muscle thin filament tropomyosin	GO:0005862	5.5E-02
Ragulator complex	GO:0071986	5.8E-02

actin filament binding	GO:0051015	9.6E-02
cell development	GO:0048468	4.8E-02
cytoskeletal protein binding	GO:0008092	9.6E-02
peptidase regulator activity	GO:0061134	9.6E-02
regulation of catalytic activity	GO:0050790	5.1E-02
supramolecular fiber organization	GO:0097435	5.8E-02
actin cytoskeleton organization	GO:0030036	5.8E-02
cellular modified amino acid metabolic process	GO:0006575	5.8E-02
response to inorganic substance	GO:0010035	5.8E-02
cellular component morphogenesis	GO:0032989	6.7E-02
regulation of cell adhesion	GO:0030155	7.0E-02
programmed cell death	GO:0012501	7.0E-02
circulatory system process	GO:0003013	7.4E-02
regulation of hemostasis	GO:1900046	7.5E-02
oligosaccharide binding	GO:0070492	1.5E-01
nucleosome	GO:0000786	1.2E-01
regulation of bone resorption	GO:0045124	9.3E-02
positive regulation of molecular function	GO:0044093	9.7E-02
transport	GO:0006810	1.0E-01
negative regulation of molecular function	GO:0044092	1.0E-01

10 Bibliography

- ABBOUD, C., BERMAN, E., COHEN, A., CORTES, J., DEANGELO, D., DEININGER, M., DEVINE, S., DRUKER, B., FATHI, A., JABBOUR, E., JAGASIA, M., KANTARJIAN, H., KHOURY, J., LANEUVILLE, P., LARSON, R., LIPTON, J., MOORE, J. O., MUGHAL, T., O'BRIEN, S., PINILLA-IBARZ, J., QUINTAS-CARDAMA, A., RADICH, J., REDDY, V., SCHIFFER, C., SHAH, N., SHAMI, P., SILVER, R. T., SNYDER, D., STONE, R., TALPAZ, M., TEFFERI, A., ETEN, R. A. V., WETZLER, M., ABRUZZESE, E., APPERLEY, J., BRECCIA, M., BYRNE, J., CERVANTES, F., CHELYSHEVA, E., CLARK, R. E., LAVALLADE, H. D., DYAGIL, I., GAMBACORTI-PASSERINI, C., GOLDMAN, J., HAZNEDAROGLU, I., HJORTH-HANSEN, H., HOLYOAKE, T., HUNTLY, B., COUTRE, P. L., LOMAIA, E., MAHON, F.-X., MARIN-COSTA, D., MARTINELLI, G., MAYER, J., MILOJKOVIC, D., OLAVARRIA, E., PORKKA, K., RICHTER, J., ROUSSELOT, P., SAGLIO, G., SAYDAM, G., STENTOF, J., TURKINA, A., VIGNERI, P., ZARITSKEY, A., AGUAYO, A., AYALA, M., BENDIT, I., BENGIO, R. M., BEST, C., BULLORSKY, E., CERVERA, E., DESOUSA, C., FANILLA, E., GOMEZ-ALMAGUER, D., HAMERSCHLAK, N., LOPEZ, J., MAGARINOS, A., MEILLON, L., MILONE, J., MOIRAGHI, B., PASQUINI, R., PAVLOVSKY, C., RUIZ-ARGUELLES, G. J., SPECTOR, N., ARTHUR, C., BROWETT, P., GRIGG, A., HU, J., HUANG, X.-J., HUGHES, T., JIANG, Q., JOOTAR, S., KIM, D.-W., MALHOTRA, H., MALHOTRA, P., MATSUMURA, I., MELO, J., OHNISHI, K., OHNO, R., et al. 2013. The price of drugs for chronic myeloid leukemia (CML) is a reflection of the unsustainable prices of cancer drugs: from the perspective of a large group of CML experts. *Blood*, 121, 4439-4442.
- ABRAHAM, S. A., HOPCROFT, L. E., CARRICK, E., DROTAR, M. E., DUNN, K., WILLIAMSON, A. J., KORFI, K., BAQUERO, P., PARK, L. E., SCOTT, M. T., PELLICANO, F., PIERCE, A., COPLAND, M., NOURSE, C., GRIMMOND, S. M., VETRIE, D., WHETTON, A. D. & HOLYOAKE, T. L. 2016. Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells. *Nature*, 534, 341-6.
- ADAMO, P. & LADOMERY, M. R. 2016. The oncogene ERG: a key factor in prostate cancer. *Oncogene*, 35, 403-14.
- ADEGOKE, O. J., BLAIR, A., SHU, X. O., SANDERSON, M., JIN, F., DOSEMECI, M., ADDY, C. L. & ZHENG, W. 2003. Occupational history and exposure and the risk of adult leukemia in Shanghai. *Ann Epidemiol*, 13, 485-94.
- ADOLFSSON, J., MÅNSSON, R., BUZA-VIDAS, N., HULTQUIST, A., LIUBA, K., JENSEN, C. T., BRYDER, D., YANG, L., BORGE, O.-J., THOREN, L. A. M., ANDERSON, K., SITNICKA, E., SASAKI, Y., SIGVARDSSON, M. & JACOBSEN, S. E. W. 2005. Identification of Flt3⁺ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential. *Cell*, 121, 295-306.
- AGIRRE, X., JIMENEZ-VELASCO, A., SAN JOSE-ENERIZ, E., GARATE, L., BANDRES, E., CORDEU, L., APARICIO, O., SAEZ, B., NAVARRO, G., VILAS-ZORNOZA, A., PEREZ-ROGER, I., GARCIA-FONCILLAS, J., TORRES, A., HEINIGER, A., CALASANZ, M. J., FORTES, P., ROMAN-GOMEZ, J. & PROSPER, F. 2008. Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34⁺ cells increases USF2-mediated cell growth. *Mol Cancer Res*, 6, 1830-40.
- AGNIHOTRI, P., ROBERTSON, N. M., UMETSU, S. E., ARAKCHEEVA, K. & WINANDY, S. 2017. Lack of Ikaros cripples expression of Foxo1 and its targets in naive T cells. *Immunology*, 152, 494-506.
- AKASHI, K., TRAVER, D., MIYAMOTO, T. & WEISSMAN, I. L. 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404, 193.

- ANDREWS, R., TAKAHASHI, M., SEGAL, G., POWELL, J., BERNSTEIN, I. & SINGER, J. 1986. The L4F3 antigen is expressed by unipotent and multipotent colony-forming cells but not by their precursors. *Blood*, 68, 1030-1035.
- ANDREWS, R. G., SINGER, J. W. & BERNSTEIN, I. D. 1989. Precursors of colony-forming cells in humans can be distinguished from colony-forming cells by expression of the CD33 and CD34 antigens and light scatter properties. *J Exp Med*, 169, 1721-31.
- ANDREWS, S. 2010. FastQC: a quality control tool for high throughput sequence data.
- ANSAR AHMED, S., GOGAL, R. M. & WALSH, J. E. 1994. A new rapid and simple non-radioactive assay to monitor and determine the proliferation of lymphocytes: an alternative to [3H]thymidine incorporation assay. *Journal of Immunological Methods*, 170, 211-224.
- APPERLEY, J. F. 2015. Chronic myeloid leukaemia. *Lancet*, 385, 1447-59.
- ARTHUR, C. K. & MA, D. D. 1993. Combined interferon alfa-2a and cytosine arabinoside as first-line treatment for chronic myeloid leukemia. *Acta Haematol*, 89 Suppl 1, 15-21.
- ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25, 25.
- BACCARANI, M., CORTES, J., PANE, F., NIEDERWIESER, D., SAGLIO, G., APPERLEY, J., CERVANTES, F., DEININGER, M., GRATWOHL, A., GUILHOT, F., HOCHHAUS, A., HOROWITZ, M., HUGHES, T., KANTARJIAN, H., LARSON, R., RADICH, J., SIMONSSON, B., SILVER, R. T., GOLDMAN, J. & HEHLMANN, R. 2009. Chronic Myeloid Leukemia: An Update of Concepts and Management Recommendations of European LeukemiaNet. *Journal of Clinical Oncology*, 27, 6041-6051.
- BACCARANI, M., DEININGER, M. W., ROSTI, G., HOCHHAUS, A., SOVERINI, S., APPERLEY, J. F., CERVANTES, F., CLARK, R. E., CORTES, J. E., GUILHOT, F., HJORTH-HANSEN, H., HUGHES, T. P., KANTARJIAN, H. M., KIM, D. W., LARSON, R. A., LIPTON, J. H., MAHON, F. X., MARTINELLI, G., MAYER, J., MULLER, M. C., NIEDERWIESER, D., PANE, F., RADICH, J. P., ROUSSELOT, P., SAGLIO, G., SAUSSELE, S., SCHIFFER, C., SILVER, R., SIMONSSON, B., STEEGMANN, J. L., GOLDMAN, J. M. & HEHLMANN, R. 2013. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. *Blood*, 122, 872-84.
- BALDUS, C. D., BURMEISTER, T., MARTUS, P., SCHWARTZ, S., GÖKBUGET, N., BLOOMFIELD, C. D., HOELZER, D., THIEL, E. & HOFMANN, W. K. 2006. High Expression of the ETS Transcription Factor ERG Predicts Adverse Outcome in Acute T-Lymphoblastic Leukemia in Adults. *Journal of Clinical Oncology*, 24, 4714-4720.
- BAQUERO, P., DAWSON, A., MUKHOPADHYAY, A., KUNTZ, E. M., MITCHELL, R., OLIVARES, O., IANNICIELLO, A., SCOTT, M. T., DUNN, K., NICASTRI, M. C., WINKLER, J. D., MICHIE, A. M., RYAN, K. M., HALSEY, C., GOTTLIEB, E., KEANEY, E. P., MURPHY, L. O., AMARAVADI, R. K., HOLYOAKE, T. L. & HELGASON, G. V. 2018. Targeting quiescent leukemic stem cells using second generation autophagy inhibitors. *Leukemia*.
- BECKER, A. J., MCCULLOCH, E. A. & TILL, J. E. 1963. Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. *Nature*, 197, 452.

- BEER, PHILIP A., KNAPP, DAVID J. H. F., KANNAN, N., MILLER, PAUL H., BABOVIC, S., BULAEVA, E., AGHAEPOUR, N., RABU, G., ROSTAMIRAD, S., SHIH, K., WEI, L. & EAVES, CONNIE J. 2014. A Dominant-Negative Isoform of IKAROS Expands Primitive Normal Human Hematopoietic Cells. *Stem Cell Reports*, 3, 841-857.
- BEINORTAS, T., TAVORIENE, I., ZVIRBLIS, T., GERBUTAVICIUS, R., JURGUTIS, M. & GRISKEVICIUS, L. 2016. Chronic myeloid leukemia incidence, survival and accessibility of tyrosine kinase inhibitors: a report from population-based Lithuanian haematological disease registry 2000-2013. *BMC Cancer*, 16, 198.
- BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57, 289-300.
- BERNITZ, J. M., KIM, H. S., MACARTHUR, B., SIEBURG, H. & MOORE, K. 2016. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell*, 167, 1296-1309.e10.
- BIERNAUX, C., LOOS, M., SELS, A., HUEZ, G. & STRYCKMANS, P. 1995. Detection of Major Bcr-Abl Gene-Expression at a Very-Low Level in Blood-Cells of Some Healthy-Individuals. *Blood*, 86, 3118-3122.
- BIGI, A., BELTRAMI, E., TRINEI, M., STENDARDO, M., PELICCI, P. G. & GIORGIO, M. 2016. Cyclophilin D counteracts P53-mediated growth arrest and promotes Ras tumorigenesis. *Oncogene*, 35, 5132-43.
- BLISS, C. I. 1939. The toxicity of poisons applied jointly. *Annals of Applied Biology*, 26, 585-615.
- BONARDI, F., FUSETTI, F., DEELEN, P., VAN GOSLIGA, D., VELLENGA, E. & SCHURINGA, J. J. 2013. A proteomics and transcriptomics approach to identify leukemic stem cell (LSC) markers. *Mol Cell Proteomics*, 12, 626-37.
- BRECCIA, M., MOLICA, M., COLAFIGLI, G., MASSARO, F., QUATTROCCHI, L., LATAGLIATA, R., MANCINI, M., DIVERIO, D., GUARINI, A., ALIMENA, G. & FOA, R. 2018. Prognostic factors associated with a stable MR4.5 achievement in chronic myeloid leukemia patients treated with imatinib. *Oncotarget*, 9, 7534-7540.
- BUTTON, K. S., IOANNIDIS, J. P. A., MOKRYSZ, C., NOSEK, B. A., FLINT, J., ROBINSON, E. S. J. & MUNAFÒ, M. R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365.
- CAI, X., GAO, L., TENG, L., GE, J., OO, ZAW M., KUMAR, ASHISH R., GILLILAND, D. G., MASON, PHILIP J., TAN, K. & SPECK, NANCY A. 2015. Runx1 Deficiency Decreases Ribosome Biogenesis and Confers Stress Resistance to Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell*, 17, 165-177.
- CARLESSO, N., FRANK, D. A. & GRIFFIN, J. D. 1996. Tyrosyl phosphorylation and DNA binding activity of signal transducers and activators of transcription (STAT) proteins in hematopoietic cell lines transformed by Bcr/Abl. *The Journal of Experimental Medicine*, 183, 811-820.
- CARREIRA, R. S., LEE, Y., GHOSHANI, M., GUSTAFSSON, A. B. & GOTTLIEB, R. A. 2010. Cyclophilin D is required for mitochondrial removal by autophagy in cardiac cells. *Autophagy*, 6, 462-72.
- CARRELHA, J., MENG, Y., KETTYLE, L. M., LUIS, T. C., NORFO, R., ALCOLEA, V., BOUKARABILA, H., GRASSO, F., GAMBARDELLA, A., GROVER, A., HÖGSTRAND, K., LORD, A. M., SANJUAN-PLA, A., WOLL, P. S., NERLOV, C. & JACOBSEN, S. E. W. 2018. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*, 554, 106.
- CARROLL, M., OHNO-JONES, S., TAMURA, S., BUCHDUNGER, E., ZIMMERMANN, J., LYDON, N. B., GILLILAND, D. G. & DRUKER, B. J. 1997.

- CGP 57148, a tyrosine kinase inhibitor, inhibits the growth of cells expressing BCR-ABL, TEL-ABL, and TEL-PDGFR fusion proteins. *Blood*, 90, 4947-52.
- CARVALHO, B. S. & IRIZARRY, R. A. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, 26, 2363-2367.
- CASHMAN, J. D., EAVES, C. J., SARRIS, A. H. & EAVES, A. C. 1998. MCP-1, not MIP-1alpha, is the endogenous chemokine that cooperates with TGF-beta to inhibit the cycling of primitive normal but not leukemic (CML) progenitors in long-term human marrow cultures. *Blood*, 92, 2338-44.
- CASTAIGNE, S., PAUTAS, C., TERRE, C., RAFFOUX, E., BORDESSOULE, D., BASTIE, J. N., LEGRAND, O., THOMAS, X., TURLURE, P., REMAN, O., DE REVEL, T., GASTAUD, L., DE GUNZBURG, N., CONTENTIN, N., HENRY, E., MAROLLEAU, J. P., ALJIJAKLI, A., ROUSSELOT, P., FENAUX, P., PREUDHOMME, C., CHEVRET, S., DOMBRET, H. & ACUTE LEUKEMIA FRENCH, A. 2012. Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet*, 379, 1508-16.
- CLAPPIER, E., AUCLERC, M. F., RAPION, J., BAKKUS, M., CAYE, A., KHEMIRI, A., GIROUX, C., HERNANDEZ, L., KABONGO, E., SAVOLA, S., LEBLANC, T., YAKOUBEN, K., PLAT, G., COSTA, V., FERSTER, A., GIRARD, S., FENNETEAU, O., CAYUELA, J. M., SIGAUX, F., DASTUGUE, N., SUCIU, S., BENOIT, Y., BERTRAND, Y., SOULIER, J. & CAVE, H. 2014. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia*, 28, 70-7.
- CLAPPIER, E., AUCLERC, M. F., RAPION, J., BAKKUS, M., CAYE, A., KHEMIRI, A., GIROUX, C., HERNANDEZ, L., KABONGO, E., SAVOLA, S., LEBLANC, T., YAKOUBEN, K., PLAT, G., COSTA, V., FERSTER, A., GIRARD, S., FENNETEAU, O., CAYUELA, J. M., SIGAUX, F., DASTUGUE, N., SUCIU, S., BENOIT, Y., BERTRAND, Y., SOULIER, J. & CAVÉ, H. 2013. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia*, 28, 70.
- CLARK, R. E., POLYDOROS, F., APPERLEY, J. F., MILOJKOVIC, D., POCOCK, C., SMITH, G., BYRNE, J. L., DE LAVALLADE, H., O'BRIEN, S. G., COFFEY, T., FORONI, L. & COPLAND, M. 2017. De-escalation of tyrosine kinase inhibitor dose in patients with chronic myeloid leukaemia with stable major molecular response (DESTINY): an interim analysis of a non-randomised, phase 2 trial. *Lancet Haematol*, 4, e310-e316.
- COPLAND, M., HAMILTON, A., ELRICK, L. J., BAIRD, J. W., ALLAN, E. K., JORDANIDES, N., BAROW, M., MOUNTFORD, J. C. & HOLYOAKE, T. L. 2006. Dasatinib (BMS-354825) targets an earlier progenitor population than imatinib in primary CML but does not eliminate the quiescent fraction. *Blood*, 107, 4532-9.
- CORBIN, A. S., O'HARE, T., GU, Z., KRAFT, I. L., EIRING, A. M., KHORASHAD, J. S., POMICTER, A. D., ZHANG, T. Y., EIDE, C. A., MANLEY, P. W., CORTES, J. E., DRUKER, B. J. & DEININGER, M. W. 2013. KIT signaling governs differential sensitivity of mature and primitive CML progenitors to tyrosine kinase inhibitors. *Cancer Res*, 73, 5775-86.
- CPMP 2001. Points to consider on switching between superiority and non-inferiority. *British Journal of Clinical Pharmacology*, 52, 223-228.
- CRAMER-MORALES, K., NIEBOROWSKA-SKORSKA, M., SCHEIBNER, K., PADGET, M., IRVINE, D. A., SLIWINSKI, T., HAAS, K., LEE, J., GENG, H., ROY, D., SLUPIANEK, A., RASSOOL, F. V., WASIK, M. A., CHILDERS, W.,

- COPLAND, M., MUSCHEN, M., CIVIN, C. I. & SKORSKI, T. 2013. Personalized synthetic lethality induced by targeting RAD52 in leukemias identified by gene mutation and expression profile. *Blood*, 122, 1293-304.
- CROCE, C. M. 2008. Molecular origins of cancer: Oncogenes and cancer. *New England Journal of Medicine*, 358, 502-511.
- CROCKER, P. R., PAULSON, J. C. & VARKI, A. 2007. Siglecs and their roles in the immune system. *Nat Rev Immunol*, 7, 255-66.
- CROSSMAN, L., MORI, M., HSIEH, Y., LANGE, T., PASCHKA, P., HARRINGTON, C., KROHN, K., NIEDERWIESER, D., HEHLMANN, R., HOCHHAUS, A., DRUKER, B. & DEININGER, M. 2005. In chronic myeloid leukemia white cells from cytogenetic responders and non-responders to imatinib have very similar gene expression signatures. *Haematologica*, 90, 459-464.
- CUNNINGHAM, F., AMODE, M. R., BARRELL, D., BEAL, K., BILLIS, K., BRENT, S., CARVALHO-SILVA, D., CLAPHAM, P., COATES, G., FITZGERALD, S., GIL, L., GIRÓN, C. G., GORDON, L., HOURLIER, T., HUNT, S. E., JANACEK, S. H., JOHNSON, N., JUETTEMANN, T., KÄHÄRI, A. K., KEENAN, S., MARTIN, F. J., MAUREL, T., MCLAREN, W., MURPHY, D. N., NAG, R., OVERDUIN, B., PARKER, A., PATRICIO, M., PERRY, E., PIGNATELLI, M., RIAT, H. S., SHEPPARD, D., TAYLOR, K., THORMANN, A., VULLO, A., WILDER, S. P., ZADISSA, A., AKEN, B. L., BIRNEY, E., HARROW, J., KINSELLA, R., MUFFATO, M., RUFFIER, M., SEARLE, S. M. J., SPUDICH, G., TREVANION, S. J., YATES, A., ZERBINO, D. R. & FLICEK, P. 2015. Ensembl 2015. *Nucleic Acids Research*, 43, D662-D669.
- CHARAF, L., MAHON, F. X., LAMRISSI-GARCIA, I., MORANVILLIER, I., BELIVEAU, F., CARDINAUD, B., DABERNAT, S., DE VERNEUIL, H., MOREAU-GAUDRY, F. & BEDEL, A. 2016. Effect of tyrosine kinase inhibitors on stemness in normal and chronic myeloid leukemia cells. *Leukemia*, 31, 65.
- CHEN, Y., WANG, H., KANTARJIAN, H. & CORTES, J. 2013. Trends in chronic myeloid leukemia incidence and survival in the United States from 1975 to 2009. *Leuk Lymphoma*, 54, 1411-7.
- CHENG, T., RODRIGUES, N., SHEN, H., YANG, Y., DOMBKOWSKI, D., SYKES, M. & SCADDEN, D. T. 2000. Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science*, 287, 1804-8.
- CHMP 2010. Guideline on the investigation of bioequivalence. In: EMA (ed.). London.
- CHOMEL, J. C., BONNET, M. L., SOREL, N., SLOMA, I., BENNACEUR-GRISCELLI, A., REA, D., LEGROS, L., MARFAING-KOKA, A., BOURHIS, J. H., AME, S., GUERCI-BRESLER, A., ROUSSELOT, P. & TURHAN, A. G. 2016. Leukemic stem cell persistence in chronic myeloid leukemia patients in deep molecular response induced by tyrosine kinase inhibitors and the impact of therapy discontinuation. *Oncotarget*.
- CHRONIC MYELOID LEUKEMIA TRIALISTS' COLLABORATIVE, G. 1997. Interferon Alfa Versus Chemotherapy for Chronic Myeloid Leukemia: A Meta-analysis of Seven Randomized Trials. *JNCI: Journal of the National Cancer Institute*, 89, 1616-1620.
- DEININGER, M. W., GOLDMAN, J. M. & MELO, J. V. 2000. The molecular biology of chronic myeloid leukemia. *Blood*, 96, 3343-56.
- DEKOTER, R. P. & SINGH, H. 2000. Regulation of B Lymphocyte and Macrophage Development by Graded Expression of PU.1. *Science*, 288, 1439-1441.
- DIJON, M., BARDIN, F., MURATI, A., BATOZ, M., CHABANNON, C. & TONNELLE, C. 2008. The role of Ikaros in human erythroid differentiation. *Blood*, 111, 1138-1146.
- DRUKER, B. J., SAWYERS, C. L., KANTARJIAN, H., RESTA, D. J., REESE, S. F., FORD, J. M., CAPDEVILLE, R. & TALPAZ, M. 2001a. Activity of a specific

- inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N Engl J Med*, 344, 1038-42.
- DRUKER, B. J., TALPAZ, M., RESTA, D. J., PENG, B., BUCHDUNGER, E., FORD, J. M., LYDON, N. B., KANTARJIAN, H., CAPDEVILLE, R., OHNO-JONES, S. & SAWYERS, C. L. 2001b. Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. *New England Journal of Medicine*, 344, 1031-1037.
- DRUKER, B. J., TAMURA, S., BUCHDUNGER, E., OHNO, S., SEGAL, G. M., FANNING, S., ZIMMERMANN, J. & LYDON, N. B. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Medicine*, 2, 561.
- DRUMMOND, M. W., HEANEY, N., KAEDA, J., NICOLINI, F. E., CLARK, R. E., WILSON, G., SHEPHERD, P., TIGHE, J., MCLINTOCK, L., HUGHES, T. & HOLYOAKE, T. L. 2009. A pilot study of continuous imatinib vs pulsed imatinib with or without G-CSF in CML patients who have achieved a complete cytogenetic response. *Leukemia*, 23, 1199-201.
- DURINCK, S., SPELLMAN, P. T., BIRNEY, E. & HUBER, W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4, 1184.
- EAVES, A. C., CASHMAN, J. D., GABOURY, L. A., KALOUSEK, D. K. & EAVES, C. J. 1986. Unregulated proliferation of primitive chronic myeloid leukemia progenitors in the presence of normal marrow adherent cells. *Proceedings of the National Academy of Sciences*, 83, 5306-5310.
- EIRING, A. M., KHORASHAD, J. S., ANDERSON, D. J., YU, F., REDWINE, H. M., MASON, C. C., REYNOLDS, K. R., CLAIR, P. M., GANTZ, K. C., ZHANG, T. Y., POMICTER, A. D., KRAFT, I. L., BOWLER, A. D., JOHNSON, K., PARTLIN, M. M., O'HARE, T. & DEININGER, M. W. 2015. beta-Catenin is required for intrinsic but not extrinsic BCR-ABL1 kinase-independent resistance to tyrosine kinase inhibitors in chronic myeloid leukemia. *Leukemia*, 29, 2328-37.
- ELISEEV, R. A., MALECKI, J., LESTER, T., ZHANG, Y., HUMPHREY, J. & GUNTER, T. E. 2009. Cyclophilin D interacts with Bcl2 and exerts an anti-apoptotic effect. *J Biol Chem*, 284, 9692-9.
- EMA 2008. Refusal assessment report for Mylotarg. London: European Medicines Agency.
- ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.
- ESSERS, M. A., OFFNER, S., BLANCO-BOSE, W. E., WAIBLER, Z., KALINKE, U., DUCHOSAL, M. A. & TRUMPP, A. 2009. IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature*, 458, 904-8.
- ESTEY, E. H. 2013. Acute myeloid leukemia: 2013 update on risk-stratification and management. *Am J Hematol*, 88, 318-27.
- FABREGAT, A., JUPE, S., MATTHEWS, L., SIDIROPOULOS, K., GILLESPIE, M., GARAPATI, P., HAW, R., JASSAL, B., KORNINGER, F., MAY, B., MILACIC, M., ROCA, C. D., ROTHFELS, K., SEVILLA, C., SHAMOVSKY, V., SHORSER, S., VARUSAI, T., VITERI, G., WEISER, J., WU, G., STEIN, L., HERMJAKOB, H. & D'EUSTACHIO, P. 2018. The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, 46, D649-D655.
- FANTA, S., SONNENBERG, M., SKORTA, I., DUYSER, J., MIETHING, C., AULITZKY, W. E. & VAN DER KUIP, H. 2008. Pharmacological inhibition of c-Abl compromises genetic stability and DNA repair in Bcr-Abl-negative cells. *Oncogene*, 27, 4380-4.

- FELDMESSER, E., OLENDER, T., KHEN, M., YANAI, I., OPHIR, R. & LANCET, D. 2006. Widespread ectopic expression of olfactory receptor genes. *BMC Genomics*, 7, 121.
- FIALKOW, P. J., GARTLER, S. M. & YOSHIDA, A. 1967. Clonal origin of chronic myelocytic leukemia in man. *Proc Natl Acad Sci U S A*, 58, 1468-71.
- FIALKOW, P. J., JACOBSON, R. J. & PAPAYANNOPOULOU, T. 1977. Chronic myelocytic leukemia: Clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage. *The American Journal of Medicine*, 63, 125-130.
- FORD, C. E., HAMERTON, J. L., BARNES, D. W. H. & LOUTIT, J. F. 1956. Cytological Identification of Radiation-Chimæras. *Nature*, 177, 452.
- GAUTIER, L., COPE, L., BOLSTAD, B. M. & IRIZARRY, R. A. 2004. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20, 307-315.
- GIUSTACCHINI, A., THONGJUEA, S., BARKAS, N., WOLL, P. S., POVINELLI, B. J., BOOTH, C. A. G., SOPP, P., NORFO, R., RODRIGUEZ-MEIRA, A., ASHLEY, N., JAMIESON, L., VYAS, P., ANDERSON, K., SEGERSTOLPE, A., QIAN, H., OLSSON-STROMBERG, U., MUSTJOKI, S., SANDBERG, R., JACOBSEN, S. E. W. & MEAD, A. J. 2017. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med*, 23, 692-702.
- GOLLNER, S., OELLERICH, T., AGRAWAL-SINGH, S., SCHENK, T., KLEIN, H. U., ROHDE, C., PABST, C., SAUER, T., LERDRUP, M., TAVOR, S., STOLZEL, F., HEROLD, S., EHNINGER, G., KOHLER, G., PAN, K. T., URLAUB, H., SERVE, H., DUGAS, M., SPIEKERMANN, K., VICK, B., JEREMIAS, I., BERDEL, W. E., HANSEN, K., ZELENT, A., WICKENHAUSER, C., MULLER, L. P., THIEDE, C. & MULLER-TIDOW, C. 2017. Loss of the histone methyltransferase EZH2 induces resistance to multiple drugs in acute myeloid leukemia. *Nat Med*, 23, 69-78.
- GRAHAM, S. M., JORGENSEN, H. G., ALLAN, E., PEARSON, C., ALCORN, M. J., RICHMOND, L. & HOLYOAKE, T. L. 2002. Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. *Blood*, 99, 319-25.
- GRATWOHL, A. 2011. The EBMT risk score. *Bone Marrow Transplantation*, 47, 749.
- GRATWOHL, A., STERN, M., BRAND, R., APPERLEY, J., BALDOMERO, H., DE WITTE, T., DINI, G., ROCHA, V., PASSWEG, J., SUREDA, A., TICHELLI, A. & NIEDERWIESER, D. 2009. Risk score for outcome after allogeneic hematopoietic stem cell transplantation: a retrospective analysis. *Cancer*, 115, 4715-26.
- GROCKOWIAK, E., LAPERROUSAZ, B., JEANPIERRE, S., VOELTZEL, T., GUYOT, B., GOBERT, S., NICOLINI, F. E. & MAGUER-SATTA, V. 2017. Immature CML cells implement a BMP autocrine loop to escape TKI treatment. *Blood*, 130, 2860-2871.
- GROWNEY, J. D., SHIGEMATSU, H., LI, Z., LEE, B. H., ADELSPERGER, J., ROWAN, R., CURLEY, D. P., KUTOK, J. L., AKASHI, K., WILLIAMS, I. R., SPECK, N. A. & GILLILAND, D. G. 2005. Loss of *Runx1* perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*, 106, 494-504.
- GUO, G., LUC, S., MARCO, E., LIN, T.-W., PENG, C., KERENYI, MARC A., BEYAZ, S., KIM, W., XU, J., DAS, PARTHA P., NEFF, T., ZOU, K., YUAN, G.-C. & ORKIN, STUART H. 2013. Mapping Cellular Hierarchy by Single-Cell Analysis of the Cell Surface Repertoire. *Cell Stem Cell*, 13, 492-505.
- GUTTERMAN, J. U. 1994. Cytokine therapeutics: lessons from interferon alpha. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 1198-1205.

- HAMILTON, A., HELGASON, G. V., SCHEMIONEK, M., ZHANG, B., MYSSINA, S., ALLAN, E. K., NICOLINI, F. E., MULLER-TIDOW, C., BHATIA, R., BRUNTON, V. G., KOSCHMIEDER, S. & HOLYOAKE, T. L. 2012. Chronic myeloid leukemia stem cells are not dependent on Bcr-Abl kinase activity for their survival. *Blood*, 119, 1501-10.
- HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell*, 144, 646-74.
- HARA, H. & OGAWA, M. 1978. Murine hemopoietic colonies in culture containing normoblasts, macrophages, and megakaryocytes. *American Journal of Hematology*, 4, 23-34.
- HARVEY, R. C., MULLIGHAN, C. G., WANG, X., DOBBIN, K. K., DAVIDSON, G. S., BEDRICK, E. J., CHEN, I. M., ATLAS, S. R., KANG, H., AR, K., WILSON, C. S., WHARTON, W., MURPHY, M., DEVIDAS, M., CARROLL, A. J., BOROWITZ, M. J., BOWMAN, W. P., DOWNING, J. R., RELLING, M., YANG, J., BHOJWANI, D., CARROLL, W. L., CAMITTA, B., REAMAN, G. H., SMITH, M., HUNGER, S. P. & WILLMAN, C. L. 2010. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*, 116, 4874-84.
- HASFORD, J., BACCARANI, M., HOFFMANN, V., GUILHOT, J., SAUSSELE, S., ROSTI, G., GUILHOT, F., PORKKA, K., OSSENKOPPELE, G., LINDOERFER, D., SIMONSSON, B., PFIRRMANN, M. & HEHLMANN, R. 2011. Predicting complete cytogenetic response and subsequent progression-free survival in 2060 patients with CML on imatinib treatment: the EUTOS score. *Blood*, 118, 686-692.
- HASFORD, J., PFIRRMANN, M., HEHLMANN, R., ALLAN, N. C., BACCARANI, M., KLUIN-NELEMANS, J. C., ALIMENA, G., STEEGMANN, J. L. & ANSARI, H. 1998. A New Prognostic Score for Survival of Patients With Chronic Myeloid Leukemia Treated With Interferon AlfaWriting Committee for the Collaborative CML Prognostic Factors Project Group. *JNCI: Journal of the National Cancer Institute*, 90, 850-859.
- HEHLMANN, R. 2012. How I treat CML blast crisis. *Blood*, 120, 737-747.
- HEHLMANN, R., HEIMPEL, H., HASFORD, J., KOLB, H. J., PRALLE, H., HOSSFELD, D. K., QUEISSER, W., LOFFLER, H., HOCHHAUS, A., HEINZE, B. & ET AL. 1994. Randomized comparison of interferon-alpha with busulfan and hydroxyurea in chronic myelogenous leukemia. The German CML Study Group. *Blood*, 84, 4064-77.
- HEINRICH, M. C., GRIFFITH, D. J., DRUKER, B. J., WAIT, C. L., OTT, K. A. & ZIGLER, A. J. 2000. Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor. *Blood*, 96, 925-932.
- HERRMANN, H., CERNY-REITERER, S., GLEIXNER, K. V., BLATT, K., HERNDLHOFFER, S., RABITSCH, W., JAGER, E., MITTERBAUER-HOHENDANNER, G., STREUBEL, B., SELZER, E., SCHWARZINGER, I., SPERR, W. R. & VALENT, P. 2012. CD34(+)/CD38(-) stem cells in chronic myeloid leukemia express Siglec-3 (CD33) and are responsive to the CD33-targeting drug gemtuzumab/ozogamicin. *Haematologica*, 97, 219-26.
- HEYSEL, R., BRILL, A. B., WOODBURY, L. A., NISHIMURA, E. T., GHOSE, T., HOSHINO, T. & YAMASAKI, M. 1960. Leukemia in Hiroshima Atomic Bomb Survivors. *Blood*, 15, 313-331.
- HIJYA, N., SCHULTZ, K. R., METZLER, M., MILLOT, F. & SUTTORP, M. 2016. Pediatric chronic myeloid leukemia is a unique disease that requires a different approach. *Blood*, 127, 392.

- HOFFMAN, R., BENZ, E. J., SILBERSTEIN, L. E., HESLOP, H., WEITZ, J. & ANASTASI, J. 2012. *Hematology: Basic Principles and Practice, Expert Consult Premium Edition - Enhanced Online Features*, Elsevier Health Sciences.
- HOGLUND, M., SANDIN, F., HELLSTROM, K., BJOREMAN, M., BJORKHOLM, M., BRUNE, M., DREIMANE, A., EKBLOM, M., LEHMANN, S., LJUNGMAN, P., MALM, C., MARKEVARN, B., MYHR-ERIKSSON, K., OHM, L., OLSSON-STROMBERG, U., SJALANDER, A., WADENVIK, H., SIMONSSON, B., STENKE, L. & RICHTER, J. 2013. Tyrosine kinase inhibitor usage, treatment outcome, and prognostic scores in CML: report from the population-based Swedish CML registry. *Blood*, 122, 1284-92.
- HOGLUND, M., SANDIN, F. & SIMONSSON, B. 2015. Epidemiology of chronic myeloid leukaemia: an update. *Ann Hematol*, 94 Suppl 2, S241-7.
- HOLTZ, M. S., SLOVAK, M. L., ZHANG, F., SAWYERS, C. L., FORMAN, S. J. & BHATIA, R. 2002. Imatinib mesylate (STI571) inhibits growth of primitive malignant progenitors in chronic myelogenous leukemia through reversal of abnormally increased proliferation. *Blood*, 99, 3792-800.
- HOLYOAKE, T., JIANG, X., EAVES, C. & EAVES, A. 1999. Isolation of a Highly Quiescent Subpopulation of Primitive Leukemic Cells in Chronic Myeloid Leukemia. *Blood*, 94, 2056-2064.
- HOLYOAKE, T. L., JIANG, X., JORGENSEN, H. G., GRAHAM, S., ALCORN, M. J., LAIRD, C., EAVES, A. C. & EAVES, C. J. 2001. Primitive quiescent leukemic cells from patients with chronic myeloid leukemia spontaneously initiate factor-independent growth in vitro in association with up-regulation of expression of interleukin-3. *Blood*, 97, 720-8.
- HOVORKOVA, L., ZALIOVA, M., VENN, N. C., BLECKMANN, K., TRKOVA, M., POTUCKOVA, E., VASKOVA, M., LINHARTOVA, J., MACHOVA POLAKOVA, K., FRONKOVA, E., MUSKOVIC, W., GILES, J. E., SHAW, P. J., CARIO, G., SUTTON, R., STARY, J., TRKA, J. & ZUNA, J. 2017. Monitoring of childhood ALL using *BCR-ABL1* genomic breakpoints identifies a subgroup with CML-like biology. *Blood*, 129, 2771-2781.
- HUGHES, T. P., SAGLIO, G., QUINTAS-CARDAMA, A., MAURO, M. J., KIM, D. W., LIPTON, J. H., BRADLEY-GARELIK, M. B., UKROPEC, J. & HOCHHAUS, A. 2015. BCR-ABL1 mutation development during first-line treatment with dasatinib or imatinib for chronic myeloid leukemia in chronic phase. *Leukemia*, 29, 1832-8.
- HUNTLY, B. J., SHIGEMATSU, H., DEGUCHI, K., LEE, B. H., MIZUNO, S., DUCLOS, N., ROWAN, R., AMARAL, S., CURLEY, D., WILLIAMS, I. R., AKASHI, K. & GILLILAND, D. G. 2004. MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell*, 6, 587-96.
- IRIZARRY, R. A., BOLSTAD, B. M., COLLIN, F., COPE, L. M., HOBBS, B. & SPEED, T. P. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*, 31, e15-e15.
- ISHIBASHI, T., YOKOTA, T., TANAKA, H., ICHII, M., SUDO, T., SATOH, Y., DOI, Y., UEDA, T., TANIMURA, A., HAMANAKA, Y., EZOE, S., SHIBAYAMA, H., ORITANI, K. & KANAKURA, Y. 2016. ESAM is a novel human hematopoietic stem cell marker associated with a subset of human leukemias. *Exp Hematol*, 44, 269-81 e1.
- JAMES, G., WITTEN, D., HASTIE, T. & TIBSHIRANI, R. 2013. *An introduction to statistical learning : with applications in R*, New York, Springer.
- JEN, E. Y., KO, C. W., LEE, J. E., DEL VALLE, P. L., AYDANIAN, A., JEWELL, C., NORSWORTHY, K. J., PRZEPIORKA, D., NIE, L., LIU, J., SHETH, C. M., SHAPIRO, M., FARRELL, A. T. & PAZDUR, R. 2018a. FDA Approval:

- Gemtuzumab ozogamicin for the treatment of adults with newly-diagnosed CD33-positive acute myeloid leukemia. *Clin Cancer Res*.
- JEN, E. Y., KO, C. W., LEE, J. E., DEL VALLE, P. L., AYDANIAN, A., JEWELL, C., NORSWORTHY, K. J., PRZEPIORKA, D., NIE, L., LIU, J., SHETH, C. M., SHAPIRO, M., FARRELL, A. T. & PAZDUR, R. 2018b. FDA Approval: Gemtuzumab Ozogamicin for the Treatment of Adults with Newly Diagnosed CD33-Positive Acute Myeloid Leukemia. *Clin Cancer Res*, 24, 3242-3246.
- JORGENSEN, H. G., ALLAN, E. K., JORDANIDES, N. E., MOUNTFORD, J. C. & HOLYOAKE, T. L. 2007. Nilotinib exerts equipotent antiproliferative effects to imatinib and does not induce apoptosis in CD34+ CML cells. *Blood*, 109, 4016-9.
- JORGENSEN, H. G., COPLAND, M., ALLAN, E. K., JIANG, X., EAVES, A., EAVES, C. & HOLYOAKE, T. L. 2006. Intermittent exposure of primitive quiescent chronic myeloid leukemia cells to granulocyte-colony stimulating factor in vitro promotes their elimination by imatinib mesylate. *Clin Cancer Res*, 12, 626-33.
- JURCIC, J. G. 2012. What happened to anti-CD33 therapy for acute myeloid leukemia? *Curr Hematol Malig Rep*, 7, 65-73.
- KAMBUROV, A., STELZL, U., LEHRACH, H. & HERWIG, R. 2013. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Research*, 41, D793-D800.
- KANEHISA, M., FURUMICHI, M., TANABE, M., SATO, Y. & MORISHIMA, K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45, D353-D361.
- KANTARJIAN, H. M., DEISSEROTH, A., KURZROCK, R., ESTROV, Z. & TALPAZ, M. 1993. Chronic myelogenous leukemia: a concise update. *Blood*, 82, 691-703.
- KANTARJIAN, H. M., O'BRIEN, S., SMITH, T. L., RIOS, M. B., CORTES, J., BERAN, M., KOLLER, C., GILES, F. J., ANDREEFF, M., KORNBLAU, S., GIRALT, S., KEATING, M. J. & TALPAZ, M. 1999. Treatment of Philadelphia chromosome-positive early chronic phase chronic myelogenous leukemia with daily doses of interferon alpha and low-dose cytarabine. *J Clin Oncol*, 17, 284-92.
- KARLEN, Y., MCNAIR, A., PERSEGUERS, S., MAZZA, C. & MERMOD, N. 2007. Statistical significance of quantitative PCR. *BMC Bioinformatics*, 8, 131.
- KAWASHIMA, H., PETRYNIAK, B., HIRAOKA, N., MITOMA, J., HUCKABY, V., NAKAYAMA, J., UCHIMURA, K., KADOMATSU, K., MURAMATSU, T., LOWE, J. B. & FUKUDA, M. 2005. N-acetylglucosamine-6-O-sulfotransferases 1 and 2 cooperatively control lymphocyte homing through L-selectin ligand biosynthesis in high endothelial venules. *Nat Immunol*, 6, 1096-104.
- KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res*, 12, 996-1006.
- KIKUCHI-TAURA, A., SOMA, T., MATSUYAMA, T., STERN, D. M. & TAGUCHI, A. 2006. A new protocol for quantifying CD34(+) cells in peripheral blood of patients with cardiovascular disease. *Tex Heart Inst J*, 33, 427-9.
- KNAPP, D., HAMMOND, C. A., HUI, T., VAN LOENHOUT, M. T. J., WANG, F., AGHAEIPOUR, N., MILLER, P. H., MOKSA, M., RABU, G. M., BEER, P. A., PELLACANI, D., HUMPHRIES, R. K., HANSEN, C., HIRST, M. & EAVES, C. J. 2018. Single-cell analysis identifies a CD33(+) subset of human cord blood cells with high regenerative potential. *Nat Cell Biol*, 20, 710-720.
- KNUDSEN, K. J., REHN, M., HASEMANN, M. S., RAPIN, N., BAGGER, F. O., OHLSSON, E., WILLER, A., FRANK, A. K., SONDERGAARD, E., JENDHOLM, J., THOREN, L., LEE, J., RAK, J., THEILGAARD-MONCH, K. & PORSE, B. T. 2015. ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev*, 29, 1915-29.

- KONDO, M., WEISSMAN, I. L. & AKASHI, K. 1997. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell*, 91, 661-672.
- KONOPKA, J. B., WATANABE, S. M. & WITTE, O. N. 1984. An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. *Cell*, 37, 1035-42.
- KOOPMAN, G., REUTELINGSPERGER, C., KUIJTEN, G., KEEHNEN, R., PALS, S. & VAN OERS, M. 1994. Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood*, 84, 1415-1420.
- KOSCHMIEDER, S., GOTTGENS, B., ZHANG, P., IWASAKI-ARAI, J., AKASHI, K., KUTOK, J. L., DAYARAM, T., GEARY, K., GREEN, A. R., TENEN, D. G. & HUETTNER, C. S. 2005. Inducible chronic phase of myeloid leukemia with expansion of hematopoietic stem cells in a transgenic model of BCR-ABL leukemogenesis. *Blood*, 105, 324-34.
- KRUEGER, F. 2015. Trim Galore!
- KULESSA, H., FRAMPTON, J. & GRAF, T. 1995. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblats, and erythroblats. *Genes & Development*, 9, 1250-1262.
- KURTOVIC-KOZARIC, A., HASIC, A., RADICH, J. P., BIJEDIC, V., NEFIC, H., EMINOVIC, I., KURTOVIC, S., COLAKOVIC, F., KOZARIC, M., VRANIC, S. & BOVAN, N. S. 2016. The reality of cancer treatment in a developing country: the effects of delayed TKI treatment on survival, cytogenetic and molecular responses in chronic myeloid leukaemia patients. *British Journal of Haematology*, 172, 420-427.
- LANDRETTE, S. F., KUO, Y. H., HENSEN, K., BARJESTE VAN WAALWIJK VAN DOORN-KHOSROVANI, S., PERRAT, P. N., VAN DE VEN, W. J., DELWEL, R. & CASTILLA, L. H. 2005. Plag1 and Plagl2 are oncogenes that induce acute myeloid leukemia in cooperation with Cbfb-MYH11. *Blood*, 105, 2900-7.
- LAURENTI, E., VARNUM-FINNEY, B., WILSON, A., FERRERO, I., BLANCO-BOSE, W. E., EHNINGER, A., KNOEPFLER, P. S., CHENG, P.-F., MACDONALD, H. R., EISENMAN, R. N., BERNSTEIN, I. D. & TRUMPP, A. 2008. Hematopoietic Stem Cell Function and Survival Depend on c-Myc and N-Myc Activity. *Cell Stem Cell*, 3, 611-624.
- LAUSEKER, M., HASFORD, J., SAUSSELE, S., KREMERS, S., KRAEMER, D., LINDEMANN, W., HEHLMANN, R. & PFIRRMANN, M. 2017. Smokers with chronic myeloid leukemia are at a higher risk of disease progression and premature death. *Cancer*, 123, 2467-2471.
- LE COUTRE, P. D., GILES, F. J., HOCHHAUS, A., APPERLEY, J. F., OSSENKOPPELE, G. J., BLAKESLEY, R., SHOU, Y., GALLAGHER, N. J., BACCARANI, M., CORTES, J. & KANTARJIAN, H. M. 2012. Nilotinib in patients with Ph⁺ chronic myeloid leukemia in accelerated phase following imatinib resistance or intolerance: 24-month follow-up results. *Leukemia*, 26, 1189-94.
- LEJNIECE, S., UDRE, I. & RIVKINA, A. 2017. Generic imatinib in the treatment of chronic myeloid leukemia: two years' experience in Latvia. *Exp Oncol*, 39, 151-154.
- LOUGHRAN, S. J., KRUSE, E. A., HACKING, D. F., DE GRAAF, C. A., HYLAND, C. D., WILLSON, T. A., HENLEY, K. J., ELLIS, S., VOSS, A. K., METCALF, D., HILTON, D. J., ALEXANDER, W. S. & KILE, B. T. 2008. The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nature Immunology*, 9, 810.
- LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.

- MAHON, F. X., REA, D., GUILHOT, J., GUILHOT, F., HUGUET, F., NICOLINI, F., LEGROS, L., CHARBONNIER, A., GUERCI, A., VARET, B., ETIENNE, G., REIFFERS, J., ROUSSELOT, P. & INTERGROUPE FRANCAIS DES LEUCEMIES MYELOIDES, C. 2010. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. *Lancet Oncol*, 11, 1029-35.
- MARCUCCI, G., BALDUS, C. D., RUPPERT, A. S., RADMACHER, M. D., MRÓZEK, K., WHITMAN, S. P., KOLITZ, J. E., EDWARDS, C. G., VARDIMAN, J. W., POWELL, B. L., BAER, M. R., MOORE, J. O., PERROTTI, D., CALIGIURI, M. A., CARROLL, A. J., LARSON, R. A., CHAPELLE, A. D. L. & BLOOMFIELD, C. D. 2005. Overexpression of the ETS-Related Gene, ERG, Predicts a Worse Outcome in Acute Myeloid Leukemia With Normal Karyotype: A Cancer and Leukemia Group B Study. *Journal of Clinical Oncology*, 23, 9234-9242.
- MARTENS, J. H. 2011. Acute myeloid leukemia: a central role for the ETS factor ERG. *Int J Biochem Cell Biol*, 43, 1413-6.
- MCWEENEY, S. K., PEMBERTON, L. C., LORIAUX, M. M., VARTANIAN, K., WILLIS, S. G., YOCHUM, G., WILMOT, B., TURPAZ, Y., PILLAI, R., DRUKER, B. J., SNEAD, J. L., MACPARTLIN, M., O'BRIEN, S. G., MELO, J. V., LANGE, T., HARRINGTON, C. A. & DEININGER, M. W. 2010. A gene expression signature of CD34+ cells to predict major cytogenetic response in chronic-phase chronic myeloid leukemia patients treated with imatinib. *Blood*, 115, 315-25.
- METROPOLIS, N. & ULAM, S. 1949. The Monte Carlo method. *J Am Stat Assoc*, 44, 335-41.
- METZKER, M. L. 2009. Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11, 31.
- MI, H., HUANG, X., MURUGANUJAN, A., TANG, H., MILLS, C., KANG, D. & THOMAS, P. D. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, 45, D183-D189.
- MORRISON, S. J. & WEISSMAN, I. L. 1994. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*, 1, 661-673.
- MUSSELMAN, J. R., BLAIR, C. K., CERHAN, J. R., NGUYEN, P., HIRSCH, B. & ROSS, J. A. 2013. Risk of adult acute and chronic myeloid leukemia with cigarette smoking and cessation. *Cancer Epidemiol*, 37, 410-6.
- NAITO, K., TAKESHITA, A., SHIGENO, K., NAKAMURA, S., FUJISAWA, S., SHINJO, K., YOSHIDA, H., OHNISHI, K., MORI, M., TERAOKAWA, S. & OHNO, R. 2000. Calicheamicin-conjugated humanized anti-CD33 monoclonal antibody (gemtuzumab zogamicin, CMA-676) shows cytotoxic effect on CD33-positive leukemia cell lines, but is inactive on P-glycoprotein-expressing sublines. *Leukemia*, 14, 1436-43.
- NAKAYAMA, H., ISHIMARU, F., KATAYAMA, Y., NAKASE, K., SEZAKI, N., TAKENAKA, K., SHINAGAWA, K., IKEDA, K., NIIYA, K. & HARADA, M. 2000. Ikaros expression in human hematopoietic lineages. *Experimental Hematology*, 28, 1232-1238.
- NERLOV, C., QUERFURTH, E., KULESSA, H. & GRAF, T. 2000. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood*, 95, 2543-2551.
- NEVIANI, P., HARB, J. G., OAKS, J. J., SANTHANAM, R., WALKER, C. J., ELLIS, J. J., FERENCHAK, G., DORRANCE, A. M., PAISIE, C. A., EIRING, A. M., MA, Y., MAO, H. C., ZHANG, B., WUNDERLICH, M., MAY, P. C., SUN, C.,

- SADDOUGHI, S. A., BIELAWSKI, J., BLUM, W., KLISOVIC, R. B., SOLT, J. A., BYRD, J. C., VOLINIA, S., CORTES, J., HUETTNER, C. S., KOSCHMIEDER, S., HOLYOAKE, T. L., DEVINE, S., CALIGIURI, M. A., CROCE, C. M., GARZON, R., OGRETMEN, B., ARLINGHAUS, R. B., CHEN, C. S., BITTMAN, R., HOKLAND, P., ROY, D. C., MILOJKOVIC, D., APPERLEY, J., GOLDMAN, J. M., REID, A., MULLOY, J. C., BHATIA, R., MARCUCCI, G. & PERROTTI, D. 2013. PP2A-activating drugs selectively eradicate TKI-resistant chronic myeloid leukemic stem cells. *J Clin Invest*, 123, 4144-57.
- NG, A. P., LOUGHRAN, S. J., METCALF, D., HYLAND, C. D., DE GRAAF, C. A., HU, Y., SMYTH, G. K., HILTON, D. J., KILE, B. T. & ALEXANDER, W. S. 2011. *Erg* is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice. *Blood*, 118, 2454-2461.
- NICOLINI, F. E., ALCAZER, V., CONY-MAKHOUL, P., HEIBLIG, M., MORISSET, S., FOSSARD, G., BIDET, A., SCHMITT, A., SOBH, M., HAYETTE, S., MAHON, F. X., DULUCQ, S. & ETIENNE, G. 2018. Long-term follow-up of de novo chronic phase chronic myelogenous leukemia patients on front-line imatinib. *Exp Hematol*.
- NOTTA, F., ZANDI, S., TAKAYAMA, N., DOBSON, S., GAN, O. I., WILSON, G., KAUFMANN, K. B., MCLEOD, J., LAURENTI, E., DUNANT, C. F., MCPHERSON, J. D., STEIN, L. D., DROR, Y. & DICK, J. E. 2016. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, 351.
- NUTT, S. L., HEAVEY, B., ROLINK, A. G. & BUSSLINGER, M. 1999. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature*, 401, 556.
- O'BRIEN, S. G., GUILHOT, F., LARSON, R. A., GATHMANN, I., BACCARANI, M., CERVANTES, F., CORNELISSEN, J. J., FISCHER, T., HOCHHAUS, A., HUGHES, T., LECHNER, K., NIELSEN, J. L., ROUSSELOT, P., REIFFERS, J., SAGLIO, G., SHEPHERD, J., SIMONSSON, B., GRATWOHL, A., GOLDMAN, J. M., KANTARJIAN, H., TAYLOR, K., VERHOEF, G., BOLTON, A. E., CAPDEVILLE, R. & DRUKER, B. J. 2003. Imatinib Compared with Interferon and Low-Dose Cytarabine for Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia. *New England Journal of Medicine*, 348, 994-1004.
- O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V. S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D. & PRUITT, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44, D733-45.
- OGAWA, M. 1993. Differentiation and proliferation of hematopoietic stem cells. *Blood*, 81, 2844-2853.
- OHNISHI, K., OHNO, R., TOMONAGA, M., KAMADA, N., ONOZAWA, K., KURAMOTO, A., DOHY, H., MIZOGUCHI, H., MIYAWAKI, S., TSUBAKI, K. & ET AL. 1995. A randomized trial comparing interferon-alpha with busulfan for newly diagnosed chronic myelogenous leukemia in chronic phase. *Blood*, 86, 906-16.

- ORKIN, S. H. 2000. Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews Genetics*, 1, 57.
- ORKIN, S. H. & ZON, L. I. 2008. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*, 132, 631-644.
- PARK, H. J., LI, J., HANNAH, R., BIDDIE, S., LEAL-CERVANTES, A. I., KIRSCHNER, K., FLORES SANTA CRUZ, D., SEXL, V., GÖTTGENS, B. & GREEN, A. R. 2016. Cytokine-induced megakaryocytic differentiation is regulated by genome-wide loss of a uSTAT transcriptional program. *The EMBO Journal*, 35, 580-594.
- PASSEGUE, E., JAMIESON, C. H., AILLES, L. E. & WEISSMAN, I. L. 2003. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci U S A*, 100 Suppl 1, 11842-9.
- PATEL, A. B., LANGE, T., POMICTER, A. D., CONLEY, C. J., HARRINGTON, C. A., REYNOLDS, K. R., KELLEY, T. W., O'HARE, T. & DEININGER, M. W. 2018. Similar expression profiles in CD34(+) cells from chronic phase chronic myeloid leukemia patients with and without deep molecular responses to nilotinib. *Oncotarget*, 9, 17889-17894.
- PAULL, T. T., ROGAKEU, E. P., YAMAZAKI, V., KIRCHGESSNER, C. U., GELLERT, M. & BONNER, W. M. 2000. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr Biol*, 10, 886-95.
- PEARCE, D. J., TAUSSIG, D. C. & BONNET, D. 2006. Implications of the expression of myeloid markers on normal and leukemic stem cells. *Cell Cycle*, 5, 271-3.
- PEDREGOSA, F., GA, #235, VAROQUAUX, L., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., #201 & DUCHESNAY, D. 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12, 2825-2830.
- PENDERGAST, A. M., QUILLIAM, L. A., CRIFE, L. D., BASSING, C. H., DAI, Z., LI, N., BATZER, A., RABUN, K. M., DER, C. J., SCHLESSINGER, J. & ET AL. 1993. BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. *Cell*, 75, 175-85.
- PRIVES, C. 1998. Signaling to p53: breaking the MDM2-p53 circuit. *Cell*, 95, 5-8.
- PRONK, C. J. H., ROSSI, D. J., MÅNSSON, R., ATTEMA, J. L., NORDDAHL, G. L., CHAN, C. K. F., SIGVARDSSON, M., WEISSMAN, I. L. & BRYDER, D. 2007. Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell*, 1, 428-442.
- PROST, S., RELOUZAT, F., SPENTCHIAN, M., OUZEGDOUH, Y., SALIBA, J., MASSONNET, G., BERESSI, J.-P., VERHOEYEN, E., RAGGUENEAU, V., MANEGLIER, B., CASTAIGNE, S., CHOMIENNE, C., CHRÉTIEN, S., ROUSSELOT, P. & LEBOULCH, P. 2015. Erosion of the chronic myeloid leukaemia stem cell pool by PPAR γ agonists. *Nature*, 525, 380.
- RADIVOYEVITCH, T., JANKOVIC, G. M., TIU, R. V., SAUNTHARARAJAH, Y., JACKSON, R. C., HLATKY, L. R., GALE, R. P. & SACHS, R. K. 2014. Sex differences in the incidence of chronic myeloid leukemia. *Radiation and environmental biophysics*, 53, 55-63.
- RADKIEWICZ, C., JOHANSSON, A. L. V., DICKMAN, P. W., LAMBE, M. & EDGREN, G. 2017. Sex differences in cancer risk and survival: A Swedish cohort study. *European Journal of Cancer*, 84, 130-140.
- REKHTMAN, N., RADPARVAR, F., EVANS, T. & SKOULTCHI, A. I. 1999. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes & Development*, 13, 1398-1411.

- RIES, L., SMITH, M., GURNEY, J., LINET, M., TAMRA, T., YOUNG, J. & BUNIN, G. 1999. Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995. . In: National Cancer Institute, vol. 99. Bethesda, MD: SEER Program; 1999:46-49.
- RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43, e47-e47.
- ROGAKOU, E. P., PILCH, D. R., ORR, A. H., IVANOVA, V. S. & BONNER, W. M. 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem*, 273, 5858-68.
- ROSS, D. M., BRANFORD, S., SEYMOUR, J. F., SCHWARER, A. P., ARTHUR, C., YEUNG, D. T., DANG, P., GOYNE, J. M., SLADER, C., FILSHIE, R. J., MILLS, A. K., MELO, J. V., WHITE, D. L., GRIGG, A. P. & HUGHES, T. P. 2013. Safety and efficacy of imatinib cessation for CML patients with stable undetectable minimal residual disease: results from the TWISTER study. *Blood*, 122, 515-22.
- ROWLEY, J. D. 1973. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, 243, 290-3.
- RUSHING, D., GOLDMAN, A., GIBBS, G., HOWE, R. & KENNEDY, B. J. 1982. Hydroxyurea versus busulfan in the treatment of chronic myelogenous leukemia. *Am J Clin Oncol*, 5, 307-13.
- SALEK-ARDAKANI, S., SMOOHA, G., DE BOER, J., SEBIRE, N. J., MORROW, M., RAINIS, L., LEE, S., WILLIAMS, O., IZRAELI, S. & BRADY, H. J. M. 2009. *ERG* Is a Megakaryocytic Oncogene. *Cancer Research*, 69, 4665-4673.
- SANJUAN-PLA, A., MACAULAY, I. C., JENSEN, C. T., WOLL, P. S., LUIS, T. C., MEAD, A., MOORE, S., CARELLA, C., MATSUOKA, S., JONES, T. B., CHOWDHURY, O., STENSON, L., LUTTEROPP, M., GREEN, J. C. A., FACCHINI, R., BOUKARABILA, H., GROVER, A., GAMBARDILLA, A., THONGJUEA, S., CARRELHA, J., TARRANT, P., ATKINSON, D., CLARK, S.-A., NERLOV, C. & JACOBSEN, S. E. W. 2013. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, 502, 232.
- SAUßELE, S. & SILVER, R. T. 2015. Management of chronic myeloid leukemia in blast crisis. *Annals of Hematology*, 94, 159-165.
- SAUVAGEAU, G., LANSDORP, P. M., EAVES, C. J., HOGGE, D. E., DRAGOWSKA, W. H., REID, D. S., LARGMAN, C., LAWRENCE, H. J. & HUMPHRIES, R. K. 1994. Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. *Proc Natl Acad Sci U S A*, 91, 12223-7.
- SAWYERS, C. L. 1999. Chronic Myeloid Leukemia. *New England Journal of Medicine*, 340, 1330-1340.
- SCOTT, M. T., KORFI, K., SAFFREY, P., HOPCROFT, L. E., KINSTRIE, R., PELLICANO, F., GUENTHER, C., GALLIPOLI, P., CRUZ, M., DUNN, K., JORGENSEN, H. G., CASSELS, J. E., HAMILTON, A., CROSSAN, A., SINCLAIR, A., HOLYOAKE, T. L. & VETRIE, D. 2016. Epigenetic Reprogramming Sensitizes CML Stem Cells to Combined EZH2 and Tyrosine Kinase Inhibition. *Cancer Discov*, 6, 1248-1257.
- SCHNEPPENHEIM, R., CASTAMAN, G., FEDERICI, A. B., KREUZ, W., MARSCHALEK, R., OLDENBURG, J., OYEN, F. & BUDDE, U. 2007. A common 253-kb deletion involving VWF and TMEM16B in German and Italian patients with severe von Willebrand disease type 3. *J Thromb Haemost*, 5, 722-8.

- SHAM, P. C. & PURCELL, S. M. 2014. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, 15, 335.
- SHTIVELMAN, E., LIFSHITZ, B., GALE, R. P. & CANAANI, E. 1985. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature*, 315, 550-4.
- SILVER, R. T., WOOLF, S. H., HEHLMANN, R., APPELBAUM, F. R., ANDERSON, J., BENNETT, C., GOLDMAN, J. M., GUILHOT, F., KANTARJIAN, H. M., LICHTIN, A. E., TALPAZ, M. & TURA, S. 1999. An evidence-based analysis of the effect of busulfan, hydroxyurea, interferon, and allogeneic bone marrow transplantation in treating the chronic phase of chronic myeloid leukemia: developed for the American Society of Hematology. *Blood*, 94, 1517-36.
- SIMINOVITCH, L., MCCULLOCH, E. A. & TILL, J. E. 1963. The distribution of colony-forming cells among spleen colonies. *Journal of Cellular and Comparative Physiology*, 62, 327-336.
- SKORSKI, T., KANAKARAJ, P., NIEBOROWSKA-SKORSKA, M., RATAJCZAK, M., WEN, S., ZON, G., GEWIRTZ, A., PERUSSIA, B. & CALABRETTA, B. 1995. Phosphatidylinositol-3 kinase activity is regulated by BCR/ABL and is required for the growth of Philadelphia chromosome-positive cells. *Blood*, 86, 726-736.
- SMYTH, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3, Article3.
- SOCOLOVSKY, M., DUSANTER-FOURT, I. & LODISH, H. F. 1997. The Prolactin Receptor and Severely Truncated Erythropoietin Receptors Support Differentiation of Erythroid Progenitors. *Journal of Biological Chemistry*, 272, 14009-14012.
- SOCOLOVSKY, M., LODISH, H. F. & DALEY, G. Q. 1998. Control of hematopoietic differentiation: Lack of specificity in signaling by cytokine receptors. *Proceedings of the National Academy of Sciences*, 95, 6573-6575.
- SOKAL, J. E., COX, E. B., BACCARANI, M., TURA, S., GOMEZ, G. A., ROBERTSON, J. E., TSO, C. Y., BRAUN, T. J., CLARKSON, B. D., CERVANTES, F. & ET AL. 1984. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. *Blood*, 63, 789-99.
- SPANGRUDE, G., HEIMFELD, S. & WEISSMAN, I. 1988. Purification and characterization of mouse hematopoietic stem cells. *Science*, 241, 58-62.
- SUDO, T., YOKOTA, T., ORITANI, K., SATOH, Y., SUGIYAMA, T., ISHIDA, T., SHIBAYAMA, H., EZOE, S., FUJITA, N., TANAKA, H., MAEDA, T., NAGASAWA, T. & KANAKURA, Y. 2012. The endothelial antigen ESAM monitors hematopoietic stem cell status between quiescence and self-renewal. *J Immunol*, 189, 200-10.
- SUGIMURA, R., JHA, D. K., HAN, A., SORIA-VALLES, C., DA ROCHA, E. L., LU, Y.-F., GOETTEL, J. A., SERRAO, E., ROWE, R. G., MALLESHAIAH, M., WONG, I., SOUSA, P., ZHU, T. N., DITADI, A., KELLER, G., ENGELMAN, A. N., SNAPPER, S. B., DOULATOV, S. & DALEY, G. Q. 2017. Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature*, 545, 432.
- SUN, J., RAMOS, A., CHAPMAN, B., JOHNNIDIS, J. B., LE, L., HO, Y.-J., KLEIN, A., HOFMANN, O. & CAMARGO, F. D. 2014. Clonal dynamics of native haematopoiesis. *Nature*, 514, 322.
- SUTHERLAND, H., EAVES, C., EAVES, A., DRAGOWSKA, W. & LANSDORP, P. 1989. Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. *Blood*, 74, 1563-1570.
- TALPAZ, M., KANTARJIAN, H. M., MCCREDIE, K. B., KEATING, M. J., TRUJILLO, J. & GUTTERMAN, J. 1987. Clinical investigation of human alpha interferon in chronic myelogenous leukemia. *Blood*, 69, 1280-8.

- TALPAZ, M., MCCREDIE, K. B., MAVLIGIT, G. M. & GUTTERMAN, J. U. 1983. Leukocyte interferon-induced myeloid cyto reduction in chronic myelogenous leukemia. *Blood*, 62, 689-92.
- TAOUDI, S., BEE, T., HILTON, A., KNEZEVIC, K., SCOTT, J., WILLSON, T. A., COLLIN, C., THOMAS, T., VOSS, A. K., KILE, B. T., ALEXANDER, W. S., PIMANDA, J. E. & HILTON, D. J. 2011. ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. *Genes & Development*, 25, 251-262.
- TERSTAPPEN, L., HUANG, S., SAFFORD, M., LANSDORP, P. & LOKEN, M. 1991. Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34⁺CD38⁻ progenitor cells. *Blood*, 77, 1218-1227.
- THE GENE ONTOLOGY CONSORTIUM 2017. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research*, 45, D331-D338.
- THOM, C. S., TRAXLER, E. A., KHANDROS, E., NICKAS, J. M., ZHOU, O. Y., LAZARUS, J. E., SILVA, A. P., PRABHU, D., YAO, Y., ARIBEANA, C., FUCHS, S. Y., MACKAY, J. P., HOLZBAUR, E. L. & WEISS, M. J. 2014. Trim58 degrades Dynein and regulates terminal erythropoiesis. *Dev Cell*, 30, 688-700.
- THOMAS, E. D., CLIFT, R. A., FEFER, A., APPELBAUM, F. R., BEATTY, P., BENSINGER, W. I., BUCKNER, C. D., CHEEVER, M. A., DEEG, H. J., DONEY, K. & ET AL. 1986. Marrow transplantation for the treatment of chronic myelogenous leukemia. *Ann Intern Med*, 104, 155-63.
- TILL, J. E. & MCCULLOCH, E. A. 1961. A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. *Radiation Research*, 14, 213-222.
- TOOFAN, P., BUSCH, C., MORRISON, H., O'BRIEN, S., JØRGENSEN, H., COPLAND, M. & WHEADON, H. 2018. Chronic myeloid leukaemia cells require the bone morphogenic protein pathway for cell cycle progression and self-renewal. *Cell Death & Disease*, 9, 927.
- TSUKAGOSHI, S. 1992. [Hydrea, an effective drug for chronic myelogenous leukemia]. *Gan To Kagaku Ryoho*, 19, 2275-80.
- TUSI, B. K., WOLOCK, S. L., WEINREB, C., HWANG, Y., HIDALGO, D., ZILIONIS, R., WAISMAN, A., HUH, J. R., KLEIN, A. M. & SOCOLOVSKY, M. 2018. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*, 555, 54.
- UCHIMURA, K., GAUGUET, J. M., SINGER, M. S., TSAY, D., KANNAGI, R., MURAMATSU, T., VON ANDRIAN, U. H. & ROSEN, S. D. 2005. A major class of L-selectin ligands is eliminated in mice deficient in two sulfotransferases expressed in high endothelial venules. *Nat Immunol*, 6, 1105-13.
- UICC 2014. Chronic Myelogenous Leukemia. *Review of Cancer Medicines on the WHO List of Essential Medicines*. WHO.
- VASEVA, A. V., MARCHENKO, N. D., JI, K., TSIRKA, S. E., HOLZMANN, S. & MOLL, U. M. 2012. p53 opens the mitochondrial permeability transition pore to trigger necrosis. *Cell*, 149, 1536-48.
- VELTEN, L., HAAS, S. F., RAFFEL, S., BLASZKIEWICZ, S., ISLAM, S., HENNIG, B. P., HIRCHE, C., LUTZ, C., BUSS, E. C., NOWAK, D., BOCH, T., HOFMANN, W.-K., HO, A. D., HUBER, W., TRUMPP, A., ESSERS, M. A. G. & STEINMETZ, LARS M. 2017. Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology*, 19, 271.
- VERMES, I., HAANEN, C., STEFFENS-NAKKEN, H. & REUTELLINGSPIERGER, C. 1995. A novel assay for apoptosis Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V. *Journal of Immunological Methods*, 184, 39-51.

- VITALE, C., ROMAGNANI, C., FALCO, M., PONTE, M., VITALE, M., MORETTA, A., BACIGALUPO, A., MORETTA, L. & MINGARI, M. C. 1999. Engagement of p75/AIRM1 or CD33 inhibits the proliferation of normal or leukemic myeloid cells. *Proc Natl Acad Sci U S A*, 96, 15091-6.
- VITALE, C., ROMAGNANI, C., PUC CETTI, A., OLIVE, D., COSTELLO, R., CHIOSSONE, L., PITTO, A., BACIGALUPO, A., MORETTA, L. & MINGARI, M. C. 2001. Surface expression and function of p75/AIRM-1 or CD33 in acute myeloid leukemias: engagement of CD33 induces apoptosis of leukemic cells. *Proc Natl Acad Sci U S A*, 98, 5764-9.
- VLAANDEREN, J., LAN, Q., KROMHOUT, H., ROTHMAN, N. & VERMEULEN, R. 2012. Occupational benzene exposure and the risk of chronic myeloid leukemia: a meta-analysis of cohort studies incorporating study quality dimensions. *Am J Ind Med*, 55, 779-85.
- WANG, W. K., QUAN, Y., FU, Q. B., LIU, Y., LIANG, Y., WU, J. W., YANG, G., LUO, C. X., OUYANG, Q. & WANG, Y. G. 2014. Dynamics between Cancer Cell Subpopulations Reveals a Model Coordinating with Both Hierarchical and Stochastic Concepts. *Plos One*, 9.
- WELLS, C. A., MOSBERGEN, R., KORN, O., CHOI, J., SEIDENMAN, N., MATIGIAN, N. A., VITALE, A. M. & SHEPHERD, J. 2013. Stemformatics: visualisation and sharing of stem cell gene expression. *Stem Cell Res*, 10, 387-95.
- WILSON, A., MURPHY, M. J., OSKARSSON, T., KALOULIS, K., BETTESS, M. D., OSER, G. M., PASCHE, A.-C., KNABENHANS, C., MACDONALD, H. R. & TRUMPP, A. 2004. c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes & Development*, 18, 2747-2763.
- WILSON, N. K., FOSTER, S. D., WANG, X., KNEZEVIC, K., SCHUTTE, J., KAIMAKIS, P., CHILARSKA, P. M., KINSTON, S., OUWEHAND, W. H., DZIERZAK, E., PIMANDA, J. E., DE BRUIJN, M. F. & GOTTGENS, B. 2010. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, 7, 532-44.
- WU, A. M., TILL, J. E., SIMINOVITCH, L. & MCCULLOCH, E. A. 1968. CYTOLOGICAL EVIDENCE FOR A RELATIONSHIP BETWEEN NORMAL HEMATOPOIETIC COLONY-FORMING CELLS AND CELLS OF THE LYMPHOID SYSTEM. *The Journal of Experimental Medicine*, 127, 455-464.
- WU, M., KWON, H. Y., RATTIS, F., BLUM, J., ZHAO, C., ASHKENAZI, R., JACKSON, T. L., GAIANO, N., OLIVER, T. & REYA, T. 2007. Imaging hematopoietic precursor division in real time. *Cell Stem Cell*, 1, 541-54.
- XIE, S., LIN, H., SUN, T. & ARLINGHAUS, R. B. 2002. Jak2 is involved in c-Myc induction by Bcr-Abl. *Oncogene*, 21, 7137-46.
- XIE, S., WANG, Y., LIU, J., SUN, T., WILSON, M. B., SMITHGALL, T. E. & ARLINGHAUS, R. B. 2001. Involvement of Jak2 tyrosine phosphorylation in Bcr-Abl transformation. *Oncogene*, 20, 6188.
- XIE, Y., KOCH, M. L., ZHANG, X., HAMBLIN, M. J., GODINHO, F. J., FUJIWARA, Y., XIE, H., KLUSMANN, J.-H., ORKIN, S. H. & LI, Z. 2017. Reduced Erg Dosage Impairs Survival of Hematopoietic Stem and Progenitor Cells. *STEM CELLS*, 35, 1773-1785.
- XIRODIMAS, D. P., SAVILLE, M. K., BOURDON, J. C., HAY, R. T. & LANE, D. P. 2004. Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. *Cell*, 118, 83-97.
- YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134.
- YOKOTA, T., ORITANI, K., BUTZ, S., KOKAME, K., KINCADE, P. W., MIYATA, T., VESTWEBER, D. & KANAKURA, Y. 2009. The endothelial antigen ESAM

- marks primitive hematopoietic progenitors throughout life in mice. *Blood*, 113, 2914-23.
- YONG, A. S., SZYDLO, R. M., GOLDMAN, J. M., APPERLEY, J. F. & MELO, J. V. 2006. Molecular profiling of CD34+ cells identifies low expression of CD7, along with high expression of proteinase 3 or elastase, as predictors of longer survival in patients with CML. *Blood*, 107, 205-12.
- ZABRISKIE, MATTHEW S., EIDE, CHRISTOPHER A., TANTRAVAH, SRINIVAS K., VELLORE, NADEEM A., ESTRADA, J., NICOLINI, FRANCK E., KHOURY, HANNA J., LARSON, RICHARD A., KONOPLEVA, M., CORTES, JORGE E., KANTARJIAN, H., JABBOUR, ELIAS J., KORNBLAU, STEVEN M., LIPTON, JEFFREY H., REA, D., STENKE, L., BARBANY, G., LANGE, T., HERNÁNDEZ-BOLUDA, J.-C., OSSENKOPPELE, GERT J., PRESS, RICHARD D., CHUAH, C., GOLDBERG, STUART L., WETZLER, M., MAHON, F.-X., ETIENNE, G., BACCARANI, M., SOVERINI, S., ROSTI, G., ROUSSELOT, P., FRIEDMAN, R., DEININGER, M., REYNOLDS, KIMBERLY R., HEATON, WILLIAM L., EIRING, ANNA M., POMICTER, ANTHONY D., KHORASHAD, JAMSHID S., KELLEY, TODD W., BARON, R., DRUKER, BRIAN J., DEININGER, MICHAEL W. & O'HARE, T. 2014. BCR-ABL1 Compound Mutations Combining Key Kinase Domain Positions Confer Clinical Resistance to Ponatinib in Ph Chromosome-Positive Leukemia. *Cancer Cell*, 26, 428-442.
- ZHANG, B., NGUYEN, L. X. T., LI, L., ZHAO, D., KUMAR, B., WU, H., LIN, A., PELLICANO, F., HOPCROFT, L., SU, Y.-L., COPLAND, M., HOLYOAKE, T. L., KUO, C. J., BHATIA, R., SNYDER, D. S., ALI, H., STEIN, A. S., BREWER, C., WANG, H., MCDONALD, T., SWIDERSKI, P., TROADEC, E., CHEN, C.-C., DORRANCE, A., PULLARKAT, V., YUAN, Y.-C., PERROTTI, D., CARLESSO, N., FORMAN, S. J., KORTYLEWSKI, M., KUO, Y.-H. & MARCUCCI, G. 2018. Bone marrow niche trafficking of miR-126 controls the self-renewal of leukemia stem cells in chronic myelogenous leukemia. *Nature Medicine*, 24, 450.
- ZHANG, P., ZHANG, X., IWAMA, A., YU, C., SMITH, K. A., MUELLER, B. U., NARRAVULA, S., TORBETT, B. E., ORKIN, S. H. & TENEN, D. G. 2000. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood*, 96, 2641-2648.
- ZHAO, X., GHAFARI, S., LODISH, H., MALASHKEVICH, V. N. & KIM, P. S. 2002. Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol*, 9, 117-20.
- ZHI, L., GAO, Y., YU, C., ZHANG, Y., ZHANG, B., YANG, J. & YAO, Z. 2016. N-Cadherin Aided in Maintaining the Characteristics of Leukemic Stem Cells. *Anat Rec (Hoboken)*, 299, 990-8.
- ZIPPER, H., BRUNNER, H., BERNHAGEN, J. & VITZTHUM, F. 2004. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research*, 32, e103-e103.