Nicht-syndromale Lippenspalten mit oder ohne Gaumenspalte und Tumorerkrankungen: Evaluation möglicher gemeinsamer genetischer Ursachen durch die Analyse von GWAS-Daten

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Hohen Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Eva Dunkhase

aus Bonn

2017

Angefertigt mit der Genehmigung

der Medizinischen Fakultät der Universität Bonn

1. Gutachter: PD Dr. Elisabeth Mangold
2. Gutachter: Prof. Dr. Andreas Jäger

Tag der Mündlichen Prüfung: 06. April 2017

Aus dem Institut für Humangenetik

Direktor: Prof. Dr. Markus M. Nöthen

Inhaltsverzeichnis

1. Deutsche Zusammenfassung

1.1 Einleitung

Orofaziale Spalten sind eine häufige angeborene Fehlbildung, die in verschiedene Subtypen eingeteilt wird. Die häufigste Form ist die nicht-syndromale „Lippenspalte mit oder ohne Gaumenspalte" (Englisch: nonsyndromic cleft lip with or without cleft palate, NSCL/P). Diese ist durch Spaltung der Oberlippe und fakultativ zusätzlich durch Spaltung des Gaumens gekennzeichnet (Mossey und Little, 2002). Die Ätiologie ist multifaktoriell, d.h. sowohl auf Umwelteinflüsse als auch auf genetische Faktoren zurückzuführen. Auch die meisten Krebserkrankungen weisen eine multifaktorielle Ätiologie auf. Molekulargenetische Studien - darunter insbesondere die in den letzten Jahren veröffentlichten genomweiten Assoziationsstudien (GWAS) - konnten Risiko-Loci sowohl für verschiedene Krebssubtypen, als auch für NSCL/P identifizieren.

Epidemiologische Studien ergaben Hinweise auf eine gemeinsame genetische Ätiologie orofazialer Spalten und spezifischer Krebserkrankungen. Eine der größten Untersuchungen in diesem Zusammenhang war die dänische Kohortenstudie von Bille et al. (2005), die eine signifikant höhere Prävalenz von Brustkrebs, Hirntumoren und Lungenkrebs bei Patienten mit verschiedenen Formen orofazialer Spalten ergab. Auch andere Studien konnten einen Zusammenhang feststellen (Christensen et al., 2004; Menezes et al., 2009; Narod et al., 1997). Blot et al. (1980), Botto et al. (2013) und Steinwachs et al. (2000) fanden hingegen keinen Hinweis auf Assoziation. Die Mehrheit der bisher veröffentlichten Untersuchungen zu einer möglichen gemeinsamen genetischen Ätiologie von orofazialen Spalten und Krebserkrankungen war rein deskriptiv. Bis heute konnten deutliche Hinweise für einen Zusammenhang zwischen Krebs und orofazialen Spalten auf molekularer Ebene nur für das *CDH1*-Gen gefunden werden (Brito et al., 2015). Die bisherigen Studien beruhten auf der Analyse von Kandidatengenen und/oder Stammbäumen. Wir verwendeten in unserer Studie einen neuen Ansatz: genomweite SNP-Daten großer Kohorten von Patienten mit sporadischer Krebserkrankung bzw. NSCL/P wurden dahingehend ausgewertet, ob gemeinsame genetische Risiko-Loci zwischen NSCL/P und denjenigen Krebsentitäten bestehen, die laut epidemiologisch basierten Arbeiten eine Häufung bei NSCL/P zeigen.

1.2 Material und Methoden

1.2.1 Suche nach GWAS zu NSCL/P-assoziierten Krebsentitäten

Wir suchten mithilfe von Pubmed (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/pubmed/) alle bis Juli 2012 veröffentlichten Studien, die einen Zusammenhang zwischen orofazialen Spalten und Krebserkrankungen untersucht hatten. Außerdem suchten wir in den Literaturverzeichnissen der so identifizierten Arbeiten nach weiteren Studien zu dieser Fragestellung. Auf dieser Grundlage erstellten wir eine Liste jener Krebsentitäten, zu denen eine Assoziation mit NSCL/P berichtet worden war (Supplementary Table S1).

Im nächsten Schritt suchten wir mithilfe des„NHGRI GWAS Online-Katalogs" (Hindorff et al., 2010) und Pubmed GWAS zu den zuvor ermittelten Krebsentitäten, für die eine Assoziation mit NSCL/P beschrieben worden war.

1.2.2 Identifizierung der Risiko-SNPs für NSCL/P und für NSCL/P-assoziierte Krebsentitäten aus der Literatur

Bis Juli 2012 waren insgesamt zwölf Risiko-Loci für NSCL/P identifiziert worden (Beaty et al., 2010; Birnbaum et al., 2009; Grant et al., 2009; Ludwig et al., 2012; Mangold et al., 2010; Rahimov et al., 2008). Von diesen Loci wurde jeweils der Top-SNP als Risiko-SNP für NSCL/P für die weiteren Analysen inunserer Studie gewählt (Table 1).

Wir erstellten außerdem eine Liste aller autosomalen SNPs, die in mindestens einer der GWAS der zuvor gewählten Krebsentitäten (vgl. Abschnitt 1.2.1) genomweite Signifikanz erreicht hatten.

1.2.3Untersuchung der genomweiten SNP-Datensätze

Für die mit Krebs assoziierten SNPs wurde die Assoziation mit NSCL/P in den Datensätzen der beiden Metaanalysen von Ludwig et al. (2012) mit i) rein europäischer (meta_Euro) und ii) kombiniert europäisch/asiatischer (meta_all) Population bestimmt.

Wir kontaktierten im Anschluss die Autoren der in Abschnitt 1.2.2 genannten Krebs-GWAS. Die Autoren wurden gebeten, die Assoziation der zwölf Top-SNPs für NSCL/P (Table 1) in ihren Datensätzen der Krebs-GWAS zu bestimmen.

Alle NSCL/P- und Krebs-SNPs, die in den jeweiligen Analyse-Datensätzen nicht vertreten waren, wurden durch einen Proxy-SNP (= Ersatz-SNP) ersetzt (vgl.

Supplementary Methods). Die *P*-Werte der Krebs-assoziierten SNPs in den NSCL/P-Metaanalysen wurden unter Verwendung eines Simulationsverfahrens für multiples Testen korrigiert. Für die Korrektur für multiples Testen der *P*-Werteder NSCL/P-assoziierten SNPs in den Krebs-GWAS wurde ein Korrekturfaktor von 384 verwendet (12 Risiko-Loci, 32 Krebs-GWAS-Datensätze).

1.3. Ergebnisse

Wir fanden zehn Publikationen, die das gleichzeitige Auftreten von NSCL/P und Krebs untersucht hatten. Hierin wurden elf primäre Krebsentitäten beschrieben: Hirntumoren, Brustkrebs, kolorektales Karzinom, Leukämie, Lymphom, Lebertumoren, Lungenkrebs, Neuroblastom, Prostatakrebs, Retinoblastom und Hautkrebs(Supplementary Table S1). Für das Retinoblastom und Lebertumoren wurden keine verwertbaren Daten in GWAS gefunden. Für die restlichen neun Krebsentitäten wurden 233 SNPs als genomweit signifikant in den jeweiligen Krebs-GWAS beschrieben. Davon wurden insgesamt 204 Krebs-SNPs in den Daten der NSCL/P-Metaanalyse auf Assoziation geprüft, da einige SNPs durch einen Proxy-SNP ersetzt bzw. wegen fehlender Proxy-SNPs ausgeschlossen werden mussten (Supplementary Tables S2.1-S2.9).

Nominale Signifikanz (P<0,05) erreichten 17 Krebs-SNPs: 14 SNPs in der meta_Euro Teilstichprobe und 12 SNPs in der meta_all Teilstichprobe (Table 2). In beiden Fällen sind dies mehr nominal signifikante SNPs als durch Zufall zu erwarten gewesen wäre (erwartete Anzahl pro Teilstichprobe: 10.2). Für acht dieser SNPs war das Krebsrisiko-Allel identisch mit dem NSCL/P-Risiko-Allel (Table 2). Der Krebs-assoziierte SNP rs6457327 auf Chromosom 6p21.33 zeigte eine Assoziation mit grenzwertiger Signifikanz im europäischen NSCL/P-Datensatz, hielt aber der Korrektur für multiples Testen nicht Stand (P_{adj} = 0,0528). Der SNP rs6457327 war in der GWAS von Skibola et al. (2009) mit dem follikulären Lymphom assoziiert. Das Risiko-Allel war in dieser GWAS wie auch im europäischen NSCL/P-Datensatz das C-Allel und somit identisch.

Die Top-SNPs der zwölf NSCL/P-Loci wurden von unseren Co-Autoren in 32 verschiedenen Krebs-Datensätzen auf Assoziation untersucht. Die Assoziation von rs13041247 auf Chromosom 20q12 mit dem Plattenepithelkarzinom der Haut blieb signifikant nach Bonferroni-Korrektur für 384 Tests (P_{adj}= 0,0018). Allerdings

unterscheiden sich die Risiko-Allele für diesen SNP bei Patienten mit Plattenepithelkarzinom der Haut und Betroffenen von NSCL/P (Table 3).
Zwei *CDH1*-SNPs (rs9929218 und rs1862748) wurden in unsere Studie einbezogen, die genomweite Signifikanz in einer GWAS zu Darmkrebs gezeigt hatten. Der SNP rs1862748 zeigte nominale Signifikanz (*P* = 0,0357) im NSCL/P-Datensatz (Table 2).
Wir übernahmen nach unserer Literatursuche 16 genomweit signifikante Krebs-SNPs aus der Region 8q24.21. Keiner dieser SNPs zeigte eine signifikante Assoziation in den NSCL/P-Datensätzen. Für den NSCL/P-SNP rs987525 aus 8q24.21 konnte in den untersuchten Krebs-Datensätzen keine Assoziation nachgewiesen werden (Table 4).

1.4 Diskussion

Das Ziel der vorliegenden Studie war es, pleiotrope Risiko-Loci für NSCL/P und diejenigen Krebsentitäten zu identifizieren, von denen ein gemeinsames Auftreten mit NSCL/P in vorherigen, überwiegend epidemiologisch basierten Studien berichtet worden war. Zwei Ansätze wurden dafür verwendet:1) Risiko-SNPs bestimmter Krebsentitäten wurden in zwei großen, genomweiten SNP-Datensätzen von NSCL/P-Patienten auf Assoziation untersucht. 2) Die Assoziation bekannter NSCL/P-Risiko-Varianten wurde mit bestimmten Krebsentitäten in entsprechenden GWAS-Datensätzen bestimmt. Dieses Vorgehen reduzierte die Anzahl der Tests und beschränkte damit die erforderliche Korrektur für multiples Testen, sodass die Wahrscheinlichkeit für die Identifikation gemeinsamer Risiko-Loci stieg.
In unseren Analysen zeigte der NSCL/P-Risiko-SNP rs13041247 auf Chromosom 20q12 auch nach Korrektur für multiples Testen eine signifikante Assoziation mit dem Plattenepithelkarzinom der Haut im Datensatz des isländischen Krebsregisters. Zu beachten ist hier allerdings, dass das NSCL/P-Risiko-Allel (T) nicht identisch ist mit dem Hautkrebsrisiko-Allel (C). rs13041247 liegt 45 kb strangabwärts des „musculo-aponeurotic fibrosarcoma oncogene homolog B“-Gen (*MAFB*), das den Transkriptionsfaktor „v-maf“ kodiert. Beaty et al. (2010) konnten mittels Tiermodellen und Sequenzierungen die Hypothese untermauern, dass *MAFB* ein Kandidatengen für die Entstehung von NSCL/P an diesem Locus ist. *MAFB* kontrolliert das „grainyhead-like transcription factor 3”-Gen *GRHL3* (Lopez-Pajares et al., 2015). Mutationen in *GRHL3* verursachen die Entstehung des Van der Woude-Syndroms, der häufigsten

syndromalen Form der Lippen-Kiefer-Gaumen-Spalten. In Hautproben von Plattenepithelkarzinomen von Mäusen und Menschen wurde eine reduzierte oder fehlende *GRHL3*-Expression beobachtet (Bhandari et al., 2013).

Der SNP rs6457327 aus 6p21.33 erreichte unter den Krebs-SNPs den niedrigsten *P*-Wert in den NSCL/P-Datensätzen. Der *P*-Wert war jedoch nach Korrektur für multiples Testen knapp nicht mehr statistisch signifikant. Dieser SNP war ursprünglich in der GWAS von Skibola et al. (2009) als Risiko-Locus für das follikuläre Lymphom aufgefallen. rs6457327 liegt 58 kb strangabwärts des „POU class 5 homeobox 1“-Gens (*POU5F1*), das auch als *OCT4* bezeichnet wird. Wang et al. (2012) zeigten, dass *OCT4* den *BMP4*-Stoffwechselweg reguliert. *BMP4* ist an der Palatogenese bei Säugern beteiligt (Zhang et al., 2002) und wurde bereits mit humanen NSCL/P in Verbindung gebracht (Chen et al., 2012; Suazo et al., 2011). *OCT4* wird in Krebszelllinien und bei verschiedenen Krebsentitäten überexprimiert (Gazouli et al., 2012; Linn et al., 2010; Yang et al., 2012). Somit ist die Hypothese plausibel, dass rs6457327 die Expression von *OCT4* reguliert, und dass hierdurch möglicherweise ein gemeinsamer zellulärer Prozess in der Onkogenese und der Entwicklung von NSCL/P gesteuert wird.

Durch die Wahl unserer Einschlusskriterien sind zwei Fehlerquellen möglich: 1) Wir achteten darauf, nur Untersuchungen zu der Assoziation des Krebsrisikos mit NSCL/P und keinem anderen Subtyp orofazialer Spalten aufzunehmen, um eventuell vorhandene genetische Heterogenität zu reduzieren. Da jedoch die Nomenklatur orofazialer Spalten nicht einheitlich angewendet wird, besteht die Möglichkeit, dass unsere Studie Krebsentitäten enthält, die mit anderen Spaltenformen als der „reinen" NSCL/P assoziiert sind. 2) Es ist möglich, dass Krebsentitäten ausgeschlossen wurden, die tatsächlich mit NSCL/P assoziiert sind, da der Zusammenhang mit NSCL/P in der jeweiligen Veröffentlichung nicht genau beschrieben war. Neuere Resequenzierungs- und Assoziationsstudien bekräftigten den Beitrag von *CDH1*-Varianten zur Entstehung von NSCL/P (Brito et al., 2015). Frebourg et al. (2006)und Kluijt et al. (2012) hatten bis Juli 2012 als einzige eine mögliche Co-Segregation von *CDH1*-bedingtem Magenkrebs und orofazialen Spalten beschrieben, die jedoch nicht auf den NSCL/P-Phänotyp beschränkt waren, sodass Magenkrebs nicht in unsere Liste der NSCL/P-assoziierten Krebsentitäten aufgenommen wurde. Dennoch waren zwei *CDH1*-SNPs (rs9929218 und rs1862748) in unserer Liste der Krebs-SNPs enthalten, da sie genomweit signifikante

Ergebnisse in einer GWAS zu Darmkrebs gezeigt hatten. rs1862748 wies nominale Signifikanz ($P = 3.57x10^{-2}$) im NSCL/P-Datensatz (meta_all) auf (Table 2).

Darüber hinaus können wir nicht ausschließen, dass NSCL/P mit anderen Krebs-Subtypen assoziiert ist, als in der vorliegenden Studie untersucht wurden, da sich viele der GWAS auf spezifische Krebs-Subtypen konzentrierten. So wurden zum Beispiel GWAS zu akuter lymphatischer Leukämie, chronisch myeloischer Leukämie, chronisch lymphatischer Leukämie, follikulärem Lymphom oder klassischem Hodgkin-Lymphom durchgeführt (Supplementary Tables S2.4 und S2.6). Diese Unterformen repräsentieren jedoch bei weitem nicht alle Leukämien und Lymphome.

Ein weiterer wichtiger Aspekt ist, dass der Altersdurchschnitt der untersuchten Personen in den deskriptiven Studien zu NSCL/P und Krebserkrankungen sehr niedrig war. Folglich ist es möglich, dass in unserer Analyse assoziierte Krebserkrankungen nicht berücksichtigt wurden, die üblicherweise erst im höheren Alter auftreten.

Aus der vorliegenden Studie zogen wir außerdem Erkenntnisse über die Region 8q24.21, die einen wichtigen Risiko-Locus für NSCL/P mit dem Top-Marker rs987525 enthält (Birnbaum et al. 2009; Mangold et al. 2011). Sechzehn der 204 Krebsrisiko-SNPs liegen in der Region 8q24.21 (Table 4). Jedoch zeigte keiner dieser SNPs eine signifikante Assoziation in den NSCL/P-Datensätzen und keine dieser Varianten steht im Kopplungsungleichgewicht mit rs987525. Dies deutet darauf hin, dass dieser Locus unterschiedliche regulatorische Regionen enthalten könnte, die für verschiedene Entwicklungsprozesse verantwortlich sind. Diese Hypothese wird durch Daten von Uslu et al. (2014) unterstützt, die zeigten, dass im benachbarten Bereich zu rs987525 ein Enhancer-Element liegt, dessen Deletion zu einer deutlichen Reduktion der *Myc*-Expression bei homozygoten Mausembryonen ausschließlich in den Gesichtsstrukturen führte.

1.5 Zusammenfassung

Zusammenfassend ist dies ist die erste Studie, die pleiotrope Risiko-Loci für NSCL/P und bestimmte Krebserkrankungen mithilfe großer, genomweiter Datensätze beschreibt. Kein Marker in unserer Studie zeigte eine signifikante Assoziation in beiden Phänotypen, jedoch ergaben sich Hinweise auf einen gemeinsamen genetischen Risikofaktor von NSCL/P und dem follikulären Lymphom in der Region 6p21.33, sowie

von NSCL/P und dem Plattenepithelkarzinom der Haut in der Region 20q12. Ob und in welchem Umfang die Ausbildung dieser Phänotypen durch eine veränderte Funktion der mutmaßlichen Kandidaten-Gene beeinflusst wird, bleibt unklar. Unsere Studie ist ein vielversprechender Ausgangspunkt für die weitere Erforschung der gemeinsamen genetischen Ätiologie von orofazialen Spalten und Krebserkrankungen.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin SC, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, and Scott AF. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 2010; 42: 525-529

Bhandari A, Gordon W, Dizon D, Hopkin AS, Gordon E, Yu Z, and Andersen B. The Grainyhead transcription factor Grhl3/Get1 suppresses miR-21 expression and tumorigenesis in skin: modulation of the miR-21 target MSH2 by RNA-binding protein DND1. Oncogene 2013; 32: 1497-1507

Bille C, Winther JF, Bautz A, Murray JC, Olsen J, and Christensen K. Cancer risk in persons with oral cleft--a population-based study of 8,093 cases. Am J Epidemiol 2005; 161: 1047-1055

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Berge SJ, Reich RH, Schiefke F, Hemprich A, Potzsch S, Steegers-Theunissen RP, Potzsch B, Moebus S, Horsthemke B, Kramer FJ, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nothen MM, and Mangold E. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 2009; 41: 473-477

Blot WJ, Stiller CA, and Wilson LM. Oral clefts and childhood cancer. Lancet 1980; 1: 722

Botto LD, Flood T, Little J, Fluchel MN, Krikov S, Feldkamp ML, Wu Y, Goedken R, Puzhankara S, and Romitti PA. Cancer risk in children and adolescents with birth defects: a population-based cohort study. PLoSOne 2013; 8: e69077

Brito LA, Yamamoto GL, Melo S, Malcher C, Ferreira SG, Figueiredo J, Alvizi L, Kobayashi GS, Naslavsky MS, Alonso N, Felix TM, Zatz M, Seruca R, and Dos Santos EP-BMR. Rare variants in the epithelial cadherin gene underlying the genetic etiology of nonsyndromic cleft lip with or without cleft palate. Hum. Mutat. 2015;36: 1029-1033

Chen Q, Wang H, Hetmanski JB, Zhang T, Ruczinski I, Schwender H, Liang KY, Fallin MD, Redett RJ, Raymond GV, Wu Chou YH, Chen PK, Yeow V, Chong SS, Cheah FS, Jabs EW, Scott AF, and Beaty TH. BMP4 was associated with NSCL/P in an Asian population. PLoSOne 2012; 7: e35347

Christensen K, Juel K, Herskind AM, and Murray JC. Long term follow up study of survival associated with cleft lip and palate at birth. BrMed J (Clin. Res. Ed). 2004; 328: 1405-1408

Frebourg T, Oliveira C, Hochain P, Karam R, Manouvrier S, Graziadio C, Vekemans M, Hartmann A, Baert-Desurmont S, Alexandre C, Lejeune Dumoulin S, Marroni C, Martin C, Castedo S, Lovett M, Winston J, Machado JC, Attie T, Jabs EW, Cai J, Pellerin P, Triboulet JP, Scotte M, Le Pessot F, Hedouin A, Carneiro F, Blayau M, and Seruca R. Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. J Med Genet 2006; 43: 138-142

Gazouli M, Roubelakis MG, Theodoropoulos GE, Papailiou J, Vaiopoulou A, Pappa KI, Nikiteas N, and Anagnou NP. OCT4 spliced variant OCT4B1 is expressed in human colorectal cancer. Mol Carcinog 2012; 51: 165-173

Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M, Frackelton EC, Otieno FG, Chiavacci RM, Nah HD, Kirschner RE, and Hakonarson H. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr 2009; 155: 909-913

Hindorff LA, Junkins HA, Hall PN, Mehta JP, and Manolio TA, 2012: A CatalogofPublished Genome-Wide Association Studies. www.genome.gov/gwastudies (Zugriffsdatum: 12.07.2012)

Kluijt I, Siemerink EJ, Ausems MG, van Os TA, de Jong D, Simoes-Correia J, van Krieken JH, Ligtenberg MJ, Figueiredo J, van Riel E, Sijmons RH, Plukker JT, van Hillegersberg R, Dekker E, Oliveira C, Cats A, and Hoogerbrugge N. CDH1-related hereditary diffuse gastric cancer syndrome: clinical variations and implications for counseling. Int JCancer 2012; 131: 367-376

Linn DE, Yang X, Sun F, Xie Y, Chen H, Jiang R, Chumsri S, Burger AM, and Qiu Y. A Role for OCT4 in Tumor Initiation of Drug-Resistant Prostate Cancer Cells. Genes Cancer 2010; 1: 908-916

Lopez-Pajares V, Qu K, Zhang J, Webster DE, Barajas BC, Siprashvili Z, Zarnegar BJ, Boxer LD, Rios EJ, Tao S, Kretz M, and Khavari PA. A LncRNA-MAF:MAFB

transcription factor network regulates epidermal differentiation. Dev Cell 2015; 32: 693-706

Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, Becker J, Herberz R, Alchawa T, Nasser E, Bohmer AC, Mattheisen M, Alblas MA, Barth S, Kluck N, Lauster C, Braumann B, Reich RH, Hemprich A, Potzsch S, Blaumeiser B, Daratsianos N, Kreusch T, Murray JC, Marazita ML, Ruczinski I, Scott AF, Beaty TH, Kramer FJ, Wienker TF, Steegers-Theunissen RP, Rubini M, Mossey PA, Hoffmann P, Lange C, Cichon S, Propping P, Knapp M, and Nothen MM. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet 2012; 44: 968-971

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, de Assis NA, Chawa TA, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, Reich RH, Hemprich A, Potzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann HE, Steegers-Theunissen RP, Kramer FJ, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey PA, Hoffmann P, and Nothen MM. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet 2010; 42: 24-26

Menezes R, Marazita ML, Goldstein McHenry T, Cooper ME, Bardi K, Brandon C, Letra A, Martin RA, and Vieira AR. AXIS inhibition protein 2, orofacial clefts and a family history of cancer. J Am Dent Assoc 2009; 140: 80-84

Mossey PA, and Little J. Epidemiology of oral clefts: An international perspective. In Wyszynski DF, ed. Cleft Lip and Palate: From Origin to Treatment. Oxford, Oxford University Press, 2002: 127-144.

Narod SA, Hawkins MM, Robertson CM, and Stiller CA. Congenital anomalies and childhood cancer in Great Britain. Am J Hum Genet 1997; 60: 474-485

National Center for Biotechnology Information, U.S. National Library of Medicine, National Institutes of Health, 2016: Pubmed. http://www.ncbi.nlm.nih.gov/pubmed/ (Zugriffsdatum: 12.07.2012)

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP, Pennacchio LA, Schutte BC, and Murray JC. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet 2008; 40: 1341-1347

Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, Iyadurai K, Becker N, Brooks-Wilson A, Curry JD, Spinelli JJ, Holly EA, Riby J, Zhang L, Nieters A, Smith MT, and Brown KM. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet 2009; 41: 873-875

Steinwachs EF, Amos C, Johnston D, Mulliken J, Stal S, and Hecht JT. Nonsyndromic cleft lip and palate is not associated with cancer or other birth defects. Am J Med Genet 2000; 90: 17-24

Suazo J, Tapia JC, Santos JL, Castro VG, Colombo A, and Blanco R. Risk variants in BMP4 promoters for nonsyndromic cleft lip/palate in a Chilean population. BMC Med Genet 2011; 12: 163

Uslu VV, Petretich M, Ruf S, Langenfeld K, Fonseca NA, Marioni JC, and Spitz F. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat. Genet. 2014; 46: 753-758

Wang Z, Oron E, Nelson B, Razis S, and Ivanova N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell 2012; 10: 440-454

Yang S, Zheng J, Ma Y, Zhu H, Xu T, Dong K, and Xiao X. Oct4 and Sox2 are overexpressed in human neuroblastoma and inhibited by chemotherapy. Oncol Rep 2012; 28: 186-192

Zhang Z, Song Y, Zhao X, Zhang X, Fermin C, and Chen Y. Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. Development 2002; 129: 4135-4146

2.Wissenschaftliche Veröffentlichung

Genomics Data 10 (2016) 22–29

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Nonsyndromic cleft lip with or without cleft palate and cancer: Evaluation of a possible common genetic background through the analysis of GWAS data

Eva Dunkhase [a], Kerstin U. Ludwig [a,b], Michael Knapp [c], Christine F. Skibola [d], Jane C. Figueiredo [e], Fay Julie Hosking [f], Eva Ellinghaus [g], Maria Teresa Landi [h], Hongxia Ma [i], Hidewaki Nakagawa [j], Jong-Won Kim [k], Jiali Han [l], Ping Yang [m], Anne C. Böhmer [b], Manuel Mattheisen [a,b,c], Markus M. Nöthen [a,b], Elisabeth Mangold [a,*]

[a] Institute of Human Genetics, University of Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany
[b] Department of Genomics, Life and Brain Center, University of Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany
[c] Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany
[d] Department of Epidemiology, University of Alabama at Birmingham, 1665 University Boulevard, Birmingham, AL 35294, USA
[e] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 1975 Zonal Ave, Los Angeles, CA 90033, USA
[f] Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, Sutton SM2 5NG, Surrey, UK
[g] Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Schittenhelmstr. 12, 24105 Kiel, Germany
[h] National Cancer Institute, NIH, DHHS, 9609 Medical Center Dr, Rockville, MD 20850, USA
[i] Department of Epidemiology, School of Public Health, Nanjing Medical University, 140 Hanzhong Rd, Gulou, Nanjing 210029, Jiangsu, China
[j] Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku,Yokohama, Kanagawa 230-0045, Japan
[k] Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, (06351) 81 Irwon-Ro Gangnam-gu. Seoul, Republic of Korea
[l] Department of Epidemiology, Richard M. Fairbanks School of Public Health, Melvin & Bren Simon Cancer Center, Indiana University, 535 Barnhill Dr, Indianapolis, IN 46202, USA
[m] Mayo Clinic, 200 First St. SW Rochester, MN 55905, USA

ARTICLE INFO

Article history:
Received 8 July 2016
Accepted 24 August 2016
Available online 26 August 2016

Keywords:
Cleft lip
Cleft palate
Genome-wide association study
Single nucleotide polymorphism
Cancer

ABSTRACT

Previous research suggests a genetic overlap between nonsyndromic cleft lip with or without cleft palate (NSCL/P) and cancer. The aim of the present study was to identify common genetic risk loci for NSCL/P and cancer entities that have been reported to co-occur with orofacial clefting. This was achieved through the investigation of large genome-wide association study datasets. Investigations of 12 NSCL/P single nucleotide polymorphisms (SNPs) in 32 cancer datasets, and 204 cancer SNPs in two NSCL/P datasets, were performed. The SNPs rs13041247 (20q12) and rs6457327 (6p21.33) showed suggestive evidence for an association with both NSCL/P and a specific cancer entity. These loci harbor genes of biological relevance to oncogenesis (MAFB and OCT4, respectively). This study is the first to characterize possible pleiotropic risk loci for NSCL/P and cancer in a systematic manner. The data represent a starting point for future research by identifying a genetic link between NSCL/P and cancer.

1. Introduction

Orofacial clefting (OFC) is a common congenital malformation comprising several subtypes. The most frequent form is nonsyndromic 'cleft lip with or without cleft palate' (NSCL/P), which is characterized by clefting of the upper lip and facultative clefting of the palate [46]. NSCL/P etiology involves both environmental and genetic factors. Genome-wide association studies (GWAS) have provided insights into the genetic background of NSCL/P through the identification of several risk loci [40,42].

Research suggests a common genetic etiology for congenital malformations - including OFC - and specific cancer entities. Miller [45] analyzed the death certificates of approximately 30,000 pediatric cancer cases in order to determine the prevalence of co-morbid congenital abnormalities. Miller reported associations between Down's syndrome and leukemia, and aniridia and Wilms' tumor. Various epidemiological study designs have since been applied to identify associations between specific cancer entities and OFC or particular clefting subtypes. One of the largest investigations to date is the Danish registry cohort study of Bille et al. [4]. The authors found a significantly higher prevalence of: breast cancer in females with cleft lip and/or cleft palate; brain cancer in females with cleft palate; and lung cancer in males with cleft lip and palate. Further studies have supported an association [13,43,47,72], whereas others have not [6,8,56].

As both cancer and OFC have a multifactorial etiology, shared risk factors might be genetic, environmental, or a combination of both. Molecular studies - in particular recent GWAS - have identified susceptibility factors for various cancer subtypes [14]. The first GWAS of cancer in the late 2000s reported a limited number of loci for the most common

* Corresponding author.
E-mail address: e.mangold@uni-bonn.de (E. Mangold).

http://dx.doi.org/10.1016/j.gdata.2016.08.017

malignancies. To date, individual studies and meta-analyses have identified approximately 50 susceptibility loci for colorectal cancer, >70 for breast cancer, and >100 for prostate cancer. Of particular interest is the finding that chromosomal region 8q24.21 contains several cancer-risk single nucleotide polymorphisms (SNPs) [25], and a major risk locus for NSCL/P [5].

The majority of published investigations into a common genetic etiology for OFC and cancer have been purely descriptive. To date, strong molecular evidence for an association between cancer and OFC has been generated for only one gene. In the respective resequencing study, a causative role for the gene *CDH1* was identified in both gastric cancer and OFC [9]. A limitation of previous studies is that they were based on the analysis of candidate genes and/or particular pedigrees. To overcome this, the present study used genome-wide SNP data from large cohorts of patients with sporadic cancers or NSCL/P. Analyses were performed to identify common genetic risk loci for NSCL/P and those cancer entities for which co-occurrence with NSCL/P has been reported. The identification of shared risk loci would provide insights into common underlying mechanisms.

2. Materials and methods

2.1. Cancer entity search strategy

In a first step, a Pubmed search was performed to locate original studies published prior to July 2012 that had investigated an association between OFC and cancer. The following search terms were used: "cleft cancer", "cleft tumor", "cleft lip cancer", "cleft lip tumor", "cleft palate cancer", "cleft palate tumor", "facial cleft cancer", "facial cleft tumor", "oral cleft cancer", "oral cleft tumor", "orofacial cleft cancer", and "orofacial cleft tumor". In the OFC literature, nomenclature is applied inconsistently, and these broad search terms were used in order to ensure that relevant studies were not overlooked due to the use of alternative morphological classification. The reference lists of these publications were then scrutinized to identify additional studies. On the basis of these search results, a list of cancer entities with a reported association with any form of OFC was compiled.

This list was then reduced to cancer entities with a reported association with NSCL/P. Studies that had investigated other OFC phenotypes (i.e., cleft palate only, syndromic OFC, or anomalies such as tooth agenesis and bifid uvula) were excluded. Data from case reports and animal models were also excluded from the analyses (Supplementary Table S1).

2.2. Cancer GWAS search strategy

In a second step, a search of the "NHGRI GWAS Online Catalog" ([29], http://www.genome.gov/gwastudies) was performed to identify GWAS of those cancer entities for which an association with NSCL/P had been reported (cancer entities of interest listed in Supplementary Table S1). Since this GWAS catalog is not exhaustive, the search was complemented using PubMed. The following search terms, followed by the respective cancer subtype, were used: "genome-wide association studies", "genome-wide association study", "gwas", "gwa".

2.3. Identification of NSCL/P risk SNPs from the literature

Prior to July 2012, three GWAS of NSCL/P were performed in European case-control samples [5,26,41], and one GWAS was performed in European and Asian trios [2]. Subsequently, two meta-analyses of GWAS data from Mangold et al. [41] and Beaty et al. [2] were performed [40]. In total, 12 loci in these studies showed genome-wide significance. These comprised one established NSCL/P risk locus (*IRF6*), and 11 novel loci. Twelve lead SNPs at these loci were chosen as NSCL/P risk SNPs for the present study (Table 1).

Table 1
NSCL/P-associated risk SNPs identified in GWAS.

SNP-ID	Allele*	Chr. region	Position (Mb)	Reference
rs560426	**G**-A	1p22.1	94.32–94.35	[2]
rs861020	**A**-G	1q32.2	208.00–208.12	[2]
rs742071	**T**-G	1p36	18.85	[40]
rs7590268	**G**-T	2p21	43.39	[40]
rs7632427	C-**T**	3p11.1	89.61	[40]
rs12543318	**C**-A	8q21.3	88.93	[40]
rs987525	**A**-C	8q24.21	129.77–130.30	[5]
rs7078160	**A**-G	10q25	118.81–118.83	[41]
rs8001641	**A**-G	13q31	79.57–79.60	[40]
rs1873147	**C**-T	15q22	61.09	[40]
rs227731	**C**-A	17q22	52.12	[41]
rs13041247	C-**T**	20q12	38.70–38.71	[2]

Chr. = chromosomal.
* Minor allele first, risk allele for NSCL/P in bold.

2.4. Identification of cancer risk SNPs from the literature

All autosomal SNPs with genome-wide significance in at least one cancer GWAS published prior to July 2012 were listed. According to the guidelines of the International HapMap Consortium [32], an SNP should be considered genome-wide significant if it achieves a *P*-value below a threshold of 5×10^{-8}. As an exception to this rule, the present study also included SNPs with a *P*-value of $>5 \times 10^{-8}$ if they had been defined as genome-wide significant in the original study.

2.5. Exploration of genome-wide SNP datasets

Data on NSCL/P-associated genetic variants were retrieved from the meta-analyses of the two largest GWAS of NSCL/P to date [40]. Ludwig et al. [40] included 497,084 SNPs, which had been genotyped in 666 complete European trios, 795 complete Asian trios, and 399 patients and 1318 controls of Central European origin. This study included 95% of all individuals available at that time with both NSCL/P and genome-wide data. For SNPs in the cancer SNP list, analyses were performed to determine association with NSCL/P in the two meta-analyses: i) European (meta_Euro); and ii) the combined European/Asian population (meta_all) datasets.

Additionally, the corresponding authors of all cancer GWAS used for SNP selection were contacted. These researchers were asked to retrieve association information from their cancer GWAS datasets for each lead SNP from the 12 NSCL/P risk loci (Table 1).

For both analyses, cancer and NSCL/P SNPs that were not represented in the respective analyzed data were replaced by a proxy SNP ($r^2 > 0.5$ in the HapMap CEU population, Supplementary methods). For the *P*-values of cancer-associated SNPs in the NSCL/P meta-analyses, correction for multiple testing was performed using a simulation procedure. This was based on 10,000 replicated samples, and involved permutation of: (i) the case and control status of individuals; and (ii) the transmitted and non-transmitted parental alleles.

For the analysis of NSCL/P-associated SNPs in the cancer GWAS, a correction factor of 384 was used (12 risk loci, 32 cancer GWAS datasets).

3. Results

The cancer entity search for studies that had analyzed the co-occurrence of orofacial anomalies and cancer identified 36 publications (Supplementary Table S1). Of these, 10 contained sufficient information to deduce that the described associations were with the NSCL/P phenotype. These 10 studies covered 11 different cancer entities, all of which were primary forms of cancer, i.e., they were not metastatic tumors. These cancer entities comprised: brain cancer [4,22,43]; breast cancer [4,43]; colorectal cancer [43]; leukemia [43,44,48,69,72]; liver cancer

Table 2
Cancer-associated SNPs with nominal significance in NSCL/P.

SNP	Risk allele cancer[a]	Risk allele NSCL/P[b]	Proxy needed?	Chr. region	P_{meta_Euro}	P_{meta_all}	Associated cancer entity	Reference
rs6457327	C	C	NO	6p21.33	1.92×10^{-4}	4.18×10^{-3}	Lymphoma (FL)	[54]
rs17505102	G	G	rs16864725	3q28	1.93×10^{-3}	2.82×10^{-2}	Leukemia (ALL)	[20]
rs3131379	n/s	A	NO	6p21.33	4.29×10^{-3}	5.69×10^{-3}	Lung cancer	[65]
rs3117582	C	C	NO	6p21.33	4.74×10^{-3}	5.46×10^{-3}	Lung cancer	[65]
rs4779584	n/s	C	NO	15q13	8.76×10^{-3}	7.76×10^{-2}	Colorectal cancer	[49]
rs10934853	A	A	NO	3q21.3	1.99×10^{-2}	1.13×10^{-3}	Prostate cancer	[27]
rs6712055	C	T	NO	2q35	2.29×10^{-2}	1.05×10^{-1}	Neuroblastoma	[10]
rs17728461	G	C	rs9614158	22q12.2	2.51×10^{-2}	1.90×10^{-1}	Lung cancer	[31]
rs4857841	A	A	NO	3q21.3	2.56×10^{-2}	1.22×10^{-3}	Prostate cancer	[27]
rs204999	A	A	NO	6p21.32	2.72×10^{-2}	2.25×10^{-1}	Lymphoma (cHL)	[15]
rs11170164	A	G	rs11170148	12q13.13	3.94×10^{-2}	3.94×10^{-2}	Skin cancer (BCC)	[55]
rs2055109	C	T	NO	3p11.2	4.24×10^{-2}	4.07×10^{-2}	Prostate cancer	[1]
rs1321311	A	A	NO	6p21.2	4.28×10^{-2}	6.41×10^{-2}	Colorectal cancer	[17]
rs10995190	G	A	NO	10q21.2	4.47×10^{-2}	4.76×10^{-2}	Breast cancer	[63]
rs4635969	C	C	NO	5p15.33	9.66×10^{-2}	3.35×10^{-2}	Lung cancer	[35]
rs17021918	C	T	NO	4q22.3	1.24×10^{-1}	1.74×10^{-2}	Prostate cancer	[19]
rs1862748	C	T	NO	16q22.1	1.50×10^{-1}	3.57×10^{-2}	Colorectal cancer	[30]

n/s = not specified; P_{meta_Euro} and P_{meta_all} = *P*-value from Likelihood ratio test in the European or European/Asian meta-analysis respectively; FL = follicular lymphoma; ALL = acute lymphoblastic leukemia; cHL = classical Hodgkin lymphoma; BCC = basal cell carcinoma.
[a] Risk allele in cancer GWAS.
[b] Risk allele in NSCL/P GWAS (in identical strand orientation).

[43]; lung cancer [4,43]; lymphoma [72]; neuroblastoma [47]; prostate cancer [43]; retinoblastoma [7]; and skin cancer [43]. The remaining 26 studies did not meet the present inclusion criteria.

Convincing GWAS results were found for nine of the 11 cancer entities. For retinoblastoma, no GWAS was found and this cancer entity was therefore excluded. The single GWAS of liver cancer had identified susceptibility variants for hepatitis B and C virus-induced hepatocellular carcinoma only [11,34,36,53,70]. Given the virus-related origin of liver cancer in these GWAS, this cancer entity was excluded from further analysis. For the remaining nine cancer entities, 233 SNPs were reported to show genome-wide significance (Supplementary Tables S2.1-S2.9). Most of these SNPs had been identified in GWAS of prostate cancer (n = 57). Only eight SNPs were derived from three GWAS of brain cancer (glioma). Sixty-six of the 233 SNPs required replacement by a proxy SNP for further analysis, and nine variants were excluded due to the lack of a proxy SNP (for details see Supplementary Tables S2.1-S2.9). In total, 204 cancer-associated SNPs were analyzed in the NSCL/P meta-analysis data.

Nominal significance ($P < 0.05$) was achieved for a total of 17 cancer SNPs: 14 cancer SNPs in the meta_Euro subsample; and 12 cancer SNPs in the meta_All subsample (Table 2). In both instances, this is more than would have been expected by chance (expected number per subsample: 10.2). For 15 of the 17 nominally significant cancer SNPs, the "cancer risk allele" was reported in the literature. For eight of

Table 3
NSCL/P associated SNPs with nominal significance in cancer GWAS.

SNP-ID (risk allele in NSCL/P)		Proxy required?	Cancer entity	Risk	Ref	P-value[a]	OR	OR type	Number of cases/controls	Sample ethnicity
rs13041247 (T)	*†	NO	skin cancer (SCC)	C	T	4.73×10^{-6}	1.23	allelic	973/>60,000	EU
rs13041247 (T)	*†	NO	skin cancer (BCC)	C	T	1.04×10^{-3}	1.10	allelic	2807/>60,000	EU
rs13041247 (T)	*†	NO	skin cancer (CM)	T	C	3.54×10^{-3}	1.16	allelic	725/>60,000	EU
rs13041247 (T)	†	NO	lymphoma (CLL)	T	C	3.31×10^{-2}	1.27	genotypic	407/296	EU[b]
rs13041247 (T)	*†	NO	lymphoma (CLL)	T	C	4.14×10^{-2}	1.25	allelic	148/>60,000	EU
rs13041247 (T)	*	NO	brain cancer (glioma)	T	C	4.18×10^{-2}	1.14	n/s	846/1310	EU[c]
rs1873147 (C)	*	NO	brain cancer (glioma)	A	G	4.50×10^{-2}	1.17	n/s	846/1310	EU[c]
rs227731 (C)	*†	NO	prostate cancer	T	G	1.40×10^{-3}	1.10	allelic	2682/>60,000	EU
rs227731 (C)	*†	NO	skin cancer (BCC)	T	G	2.35×10^{-3}	1.09	allelic	2807/>60,000	EU
rs227731 (C)	*†	NO	skin cancer (CM)	T	G	1.83×10^{-2}	1.14	allelic	725/>60,000	EU
rs227731 (C)	*	NO	skin cancer (BCC)	T	G	3.29×10^{-2}	1.08	allelic	2045/6013	EU
rs560426 (G)	*	NO	prostate cancer	T	C	5.83×10^{-4}	1.19	n/s	1583/4944	AS
rs560426 (G)	*†	NO	colorectal cancer	C	T	1.85×10^{-3}	1.06	allelic	12,620/15,110	EU
rs560426 (G)	*	NO	lymphoma (DLBCL)	C	T	4.23×10^{-2}	1.23	allelic	256/747	EU
rs7078160 (A)	*†	NO	skin cancer (BCC)	G	A	1.62×10^{-3}	1.14	allelic	2807/>60,000	EU
rs7078160 (A)	*	NO	brain cancer (glioma)	A	G	3.12×10^{-2}	1.16	n/s	1247/2236	EU[d]
rs7078160 (A)	*†	NO	skin cancer (CM)	G	A	3.15×10^{-2}	1.16	allelic	725/>60,000	EU
rs7632427 (T)	*	NO	skin cancer (SCC)	C	T	1.46×10^{-2}	1.12	allelic	973/>60,000	EU
rs7632427 (T)	†	NO	brain cancer (glioma)	T	C	4.84×10^{-2}	1.10	additive	2331/3077	AS
rs861020 (A)	‡	rs1962735	leukemia (ALL)	G	A	3.13×10^{-2}	1.28	n/s	1696/3535	EU
rs861020 (A)	*	NO	lymphoma (FL)	G	A	4.97×10^{-2}	1.33	allelic	213/750	EU
rs987525 (A)	*	NO	brain cancer (glioma)	A	C	1.15×10^{-2}	1.20	n/s	846/1310	EU[c]
rs987525 (A)	*	NO	brain cancer (glioma)	A	C	3.65×10^{-2}	1.14	n/s	1423/1190	EU[e]

Risk = Risk allele in cancer GWAS; Ref = Reference allele in cancer GWAS; OR = Odds Ratio; * SNP genotyped; † SNP imputed; ‡ no data available for SNP; SCC = squamous cell carcinoma; BCC = basal cell carcinoma; CM = cutaneous melanoma; CLL = chronic lymphocytic leukemia; DLBCL = Diffuse large B-cell lymphoma; FL = follicular lymphoma; EU = European ethnicity; AS = Asian ethnicity; n/s = not specified.
[a] Association P-value from cancer GWAS.
[b] 99% are known or assumed to be White and Not Hispanic.
[c] German subgroup.
[d] US subgroup.
[e] French subgroup.

Table 4
Cancer risk SNPs at 8q24.21 and distance from NSCL/P risk SNP rs987525.

Cancer entity	SNP-ID	Risk allele[a]	OR[b]	Position	Distance from rs987525 (kb)	Reference
Breast cancer	rs13281615	C	1.08	128,424,800	−1591	[18]
Colorectal cancer	rs6983267	G	1.21	128,476,625	−1539	[62]
Colorectal cancer	rs10505477	n/s	1.12	128,482,487	−1533	[62]
Colorectal cancer	rs7014346	A	1.19	128,493,974	−1521	[60]
Glioma	rs4295627	G	1.36	130,754,639	739	[52]
Lymphoma (cHL)	rs2608053	G	1.20	129,261,453	−754	[21]
Lymphoma (cHL)	rs2019960	G	1.33	129,145,014	−870	[21]
Lymphoma (CLL)	rs2456449	G	1.26	128,262,163	−1753	[16]
Lymphoma (CLL)	rs2466024	A	1.20	128,257,201	−1758	[16]
Prostate cancer	rs1447295	A	1.60	128,554,220	−1461	[28]
Prostate cancer	rs16901979	A	1.79	128,194,098	−1821	[28]
Prostate cancer	rs6983267	G	1.27	128,482,487	−1533	[68]
Prostate cancer	rs7837688	T	1.47	128,608,542	−1407	[68]
Prostate cancer	rs4242382	A	n/s	128,586,755	−1429	[61]
Prostate cancer	rs16902094	G	1.21	128,389,528	−1626	[27]
Prostate cancer	rs445114	T	1.14	128,410,090	−1605	[27]
Prostate cancer	rs16902104	T	1.21	128,392,363	−1623	[27]

n/s = not specified; cHL = classical Hodgkin's lymphoma; CLL = chronic lymphocytic leukemia.
[a] Risk allele in cancer GWAS.
[b] Allelic odds ratio in cancer GWAS.

these SNPs, the "cancer risk allele" was identical to the "NSCL/P risk allele" (Table 2). For the cancer-associated SNP rs6457327 on chromosome 6p21.33, a borderline association was reported in the European NSCL/P dataset. However, this became non-significant following correction for multiple testing ($P_{adj} = 0.0528$). In the original report, the SNP rs6457327 was associated with follicular lymphoma [54]. In the original cancer GWAS, risk was conferred by the C allele at this SNP, which is identical to the risk allele in the NSCL/P meta-analyses datasets.

Analyses were then performed to identify associations with the lead SNPs of the 12 NSCL/P loci in 32 different cancer sample datasets. Eight SNPs achieved nominal significance in at least one sample dataset. The association of rs13041247 at chromosome 20q12 with squamous cell cancer of the skin ($P_{adj} = 4.73 \times 10^{-6} \times 384 = 0.0018$, data extracted from the Icelandic Cancer Registry) remained significant after conservative Bonferroni correction for 384 tests. However, the risk allele for this SNP reported in squamous cell cancer of the skin differs from that found in the NSCL/P patients (Table 3).

Two *CDH1* SNPs (rs9929218 and rs1862748) were included in the present study as they had shown genome-wide significant results in a GWAS of colorectal cancer [30]. The SNP rs1862748 showed nominal significance ($P = 3.57 \times 10^{-2}$) in the present NSCL/P dataset (Table 2).

The literature search for cancer risk SNPs identified 16 SNPs in the region of 8q24.21. This region also contains a key susceptibility locus for NSCL/P, with the top marker being rs987525. None of the cancer risk SNPs showed significant association in the NSCL/P datasets, and no association with rs987525 was found in any of the investigated cancer datasets (Table 4).

4. Discussion

The aim of the present study was to identify common genetic risk loci for NSCL/P and those cancer entities that have been reported to co-occur with NSCL/P in descriptive studies. Two approaches were used. First, conclusively identified cancer susceptibility variants were analyzed in a large genome-wide SNP dataset of NSCL/P patients. Second, known NSCL/P risk loci were analyzed in GWAS data for specific cancer entities. Analysis of only a subset of candidate SNPs, i.e., those identified as being genome-wide significant for one trait, reduced the number of tests and thus the requirement for correction, thereby increasing the chances of identifying common risk loci.

In principle, an overlapping genetic contribution to both traits could also be quantified using a genome wide polygenic score approach [51]. However, for this etiological overlap to be apparent in a polygenic score, a specific cancer entity and NSCL/P would have to share a large number of genetic risk factors. Given the weak associations reported for cancer entities and NSCL/P in previous epidemiological studies, the presence of a large number of shared genetic risk factors cannot be assumed. In addition, since the polygenic score allows no conclusions to be drawn concerning a particular gene, this approach would allow no conclusions concerning a common biological pathway for NSCL/P and any specific cancer entity. We therefore considered the present methodology to be the more appropriate approach to our research question.

One association withstood correction for multiple testing. The NSCL/P-associated risk SNP rs13041247 at chromosome 20q12 showed genome-wide significant association in the dataset of the Icelandic cancer registry for squamous cell carcinoma of the skin. However, the NSCL/P risk allele (T) is not identical to the skin cancer risk allele (C). The SNP rs13041247 maps 45 kb downstream of the musculoaponeurotic fibrosarcoma oncogene homolog B (*MAFB*) gene, which encodes the v-maf transcription factor. In the GWAS of NSCL/P conducted by Beaty et al. [2], which was the first to describe association of this variant with NSCL/P, sequencing of conserved elements within the 3′ region and the coding region of MAFB in NSCL/P cases and controls from Iowa and the Philippines revealed an overrepresentation of a rare missense variant (His131Gln). The contribution of this variant to NSCL/P awaits elucidation [2]. Animal studies have provided additional support for the hypothesis that *MAFB* is a candidate gene for NSCL/P at this locus by showing that its homolog in rodents is expressed in craniofacial structures during embryogenesis [2]. Recently, Lopez-Pajares et al. [39] demonstrated that MAF and MAFB control the expression of the transcription factor genes *GRHL3*, *ZNF750*, *PRDM1*, and *KLF4*. Together, these genes form a network that is essential for epidermal differentiation. Previous studies have demonstrated that the grainyhead-like transcription factor 3 gene *GRHL3* causes the autosomal dominant Van der Woude syndrome [50], which is the most common syndromic form of cleft lip and palate. Furthermore, Bhandari et al. [3] observed a marked reduction or absence of *GRHL3* expression in squamous cell skin carcinoma samples from mice and humans. A recent study identified dominant negative KLF4 variants in patients with NSCL/P [38].

Of the cancer associated SNPs, the most significant *P*-value in the NSCL/P meta-analyses datasets was found for rs6457327 at 6p21.33, although this fell short of significance after correction for multiple testing. This SNP was originally reported by Skibola et al. [54] as a risk locus for follicular lymphoma, and maps 58 kb downstream of the POU class 5 homeobox 1 (*POU5F1*) gene, also known as *OCT4*. This gene encodes a transcription factor with an important role in embryonic development, in particular during early embryogenesis, and which is necessary for maintaining embryonic stem cell pluripotency [57]. Wang et al. [66]

showed that OCT4 regulates and interacts with the BMP4 pathway in specifying different developmental fates in human embryonic stem cells. Notably, the BMP4 pathway is involved in mammalian palatogenesis [71], and mutations in *BMP4* have been associated with human NSCL/P [12,58,59]. Research has shown that *OCT4* is overexpressed in cancer cell lines and in diverse cancer entities [24,37,67], suggesting that aberrant transcriptional regulation of *OCT4* might be a mechanism in cancer susceptibility. Thus, a plausible hypothesis is that rs6457327 regulates *OCT4* expression, and that this regulation is a possible common process in oncogenesis and the development of NSCL/P.

An important consideration in interpreting the present results is that the publication search adhered to very strict inclusion criteria, which might have introduced two sources of bias. First, we concentrated on investigating cancer risk association with NSCL/P and no other OFC subtype in order to reduce any existing genetic heterogeneity. However, as OFC nomenclature is applied inconsistently, we cannot exclude the possibility that our investigation included cancer entities that were associated with forms of OFC other than "pure" NSCL/P. Second, we may have rejected genuinely associated cancer entities due to a non-precise description of a possible association with NSCL/P in the respective publication. Frebourg et al. [23] and Kluijt et al. [33] described a possible co-segregation of OFC and *CDH1*-related gastric cancer, and recent resequencing and association studies strongly support a contribution to NSCL/P of predominantly rare and moderately penetrant *CDH1* variants [9]. As this co-occurrence was not precisely related to NSCL/P in the initial studies, and no other study of this cancer entity had been published by July 2012, gastric cancer was not included in the present cancer entity list. Nonetheless, two *CDH1* SNPs (rs9929218 and rs1862748) were included, as they had shown genome-wide significant results in a GWAS of colorectal cancer [30], and rs1862748 showed nominal significance ($P = 3.57 \times 10^{-2}$) in the present NSCL/P dataset (Table 2).

Another important consideration is that most of the individuals investigated in the descriptive studies had a relatively low mean- and median age. Bille et al. [4] were the only authors to investigate the co-occurrence of OFC and cancer in adults (maximum age: 62 years). Consequently, the present analyses may have failed to consider cancers that are more frequent in later life.

In addition, we cannot exclude the possibility that NSCL/P is associated with cancer-subtypes other than those considered in the present study. Many of the GWAS concentrated on specific cancer subtypes. For example, studies of leukemia and lymphoma were conducted using case cohorts of acute lymphoblastic leukemia, chronic myelogenous leukemia, chronic lymphocytic leukemia, follicular lymphoma, and Hodgkin's lymphoma (Supplementary Tables S2.4 and S2.6), which do not represent all forms of leukemia and lymphoma. This also applies to the skin cancer studies, since most of the GWAS of skin cancer were performed for the most frequent skin cancer subtypes, such as cutaneous melanoma, basal cell carcinoma, and squamous cell carcinoma.

An interesting result of the present study is the finding for the 8q24.21 region. This contains a key susceptibility locus for NSCL/P, and although it was initially identified in a small GWAS, the genetic effect was very pronounced [5]. The finding has since been confirmed in numerous GWAS and targeted replication studies [42]. The top marker, rs987525, maps to an intergenic region, which may contain remote cis-acting enhancers that control expression of the well known proto-oncogene *Myc* in the developing murine facial prominences [64]. Sixteen of the present cancer-risk SNPs are located in the 8q24.21 region (Table 4). However, none of these SNPs showed a statistically significant association in the NSCL/P data-sets. Furthermore, none of these cancer risk variants is in linkage disequilibrium with rs987525, which suggests that this locus might contain distinct regulatory regions that are responsible for different developmental processes. This hypothesis is supported by data from Uslu et al. [64], who showed that a distinct region adjacent to rs987525 contained a specific facial enhancer element. Deletion of this medial-nasal enhancer resulted in a pronounced reduction in *myc* expression in the facial tissues of homozygous murine embryos but not in other embryonic tissues. Therefore it is possible that the 8q24 region contains enhancers that control the expression of *Myc* in either facial development or cancer but not in both.

In summary, the present study is the first to characterize possible pleiotropic risk loci for NSCL/P and cancer using large genome-wide datasets. Suggestive evidence for a common genetic background was found for NSCL/P and follicular lymphoma at 6p21.33, and for NSCL/P and squamous cell carcinoma of the skin at 20q12. Whether, and to what extent, the development of these phenotypes is influenced by an altered function of the putative candidate genes *OCT4* and *MAFB* at these loci remains unclear. No marker in the present study showed pronounced effects on both phenotypes. Although inconclusive at the single marker level, the present data represent a starting point for further research into the common genetic etiology of OFC and cancer.

Conflicts of interest

None

GECCO funding

GECCO: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01 CA059045).

ASTERISK: a Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

COLO2&3: National Institutes of Health (R01 CA60987).

CCFR: This work was supported by grant UM1 CA167551 from the National Cancer Institute and through cooperative agreements with the following CCFR centers:

Australasian Colorectal Cancer Family Registry (U01 CA074778 and U01/U24 CA097735).

Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01/U24 CA074800).

Ontario Familial Colorectal Cancer Registry (U01/U24 CA074783).

Seattle Colorectal Cancer Family Registry (U01/U24 CA074794).

University of Hawaii Colorectal Cancer Family Registry (U01/U24 CA074806).

USC Consortium Colorectal Cancer Family Registry (U01/U24 CA074799).

The Colon CFR GWAS was supported by funding from the National Cancer Institute, National Institutes of Health (U01 CA122839 and R01 CA143237 to Graham Casey). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CCFR.

DACHS: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814).

DALS: National Institutes of Health (R01 CA48998 to M. L. Slattery);

HPFS is supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, R01 CA151993 and P50 CA127003), NHS by the National Institutes of Health (UM1 CA186107, R01 CA137178, P01 CA87969, R01 CA151993 and P50 CA127003,) and PHS by the National Institutes of Health (R01 CA042182).

MEC: National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA63464).

OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. Additional funding toward genetic

analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager, M et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007 May;39(5):645-9), Colon CGEMS pancreatic cancer scan (PanScan) (Amundadottir, L et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009 Sep;41(9):986-90, and Petersen, GM et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010 Mar;42(3):224-8), and the Lung Cancer and Smoking study (Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009 Nov;85(5):679-91). The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping.

PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb).

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

GECCO acknowledgements

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible. The authors acknowledge Dave Duggan and team members at TGEN (Translational Genomics Research Institute), the Broad Institute, and the Génome Québec Innovation Center for genotyping DNA samples of cases and controls, and for scientific input for GECCO.

HPFS, NHS and PHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

PLCO: The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs. Bill Kopp and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Acknowledgements

We thank all affected individuals and their families for their participation in this study, as well as the German support group for persons with cleft lip and/or palate (Deutsche Selbsthilfevereinigung für Lippen-Gaumen-Fehlbildungen e.V.).The study was supported by the Deutsche Forschungsgemeinschaft (FOR 423 and individual grants MA 2546/3-1, KR 1912/7-1, NO 246/6-1 and WI 1555/5-1). In particular, we thank the following researchers for the provision of data: deCODE Genetics (Reykjavik, Iceland); Victor Enciso (The Institute of Cancer Research, Sutton, Surrey, UK); Susan L. Slager (Mayo Clinic, Rochester, MN, USA); and Peter Kraft (Harvard School of Public Health, Boston, MA, USA). The study was also supported by Richard S. Houlston (The Institute of Cancer Research, Sutton, Surrey, UK); Noralane M. Lindor (Mayo Clinic, Department of Health Sciences Research, Mayo Clinic Arizona, USA); and Zhibin Hu and Hongbing Shen of Nanjing Medical University.

Christine F Skibola was supported by the National Institutes of Health RO1CA1046282 and RO1CA154643.

Ping Yang received support from grants NCI-CA77118 and CA80127, and from the Mayo Foundation.

Hidewaki Nakagawa was supported by BioBank Japan.

Jong-Won Kim was supported by the Korean Health Technology R&D Project (A120030).

The Environment and Genetics in Lung Cancer Etiology (EAGLE), Prostate, Lung, Colon, Ovary Screening Trial (PLCO), and Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) studies were supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute (NCI), Division of Cancer Epidemiology and Genetics. ATBC was also supported by U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI. PLCO was also supported by individual contracts from the NCI to the University of Colorado Denver (NO1-CN-25514), Georgetown University (NO1-CN-25522), the Pacific Health Research Institute (NO1-CN-25515), the Henry Ford Health System (NO1-CN-25512), the University of Minnesota, (NO1-CN-25513), Washington University (NO1-CN-25516), the University of Pittsburgh (NO1-CN-25511), the University of Utah (NO1-CN-25524), the Marshfield Clinic Research Foundation (NO1-CN-25518), the University of Alabama at Birmingham (NO1-CN-75022), Westat, Inc. (NO1-CN-25476), and the University of California, Los Angeles (NO1-CN-25404). The Cancer Prevention Study-II (CPS-II) Nutrition Cohort was supported by the American Cancer Society. The NIH Genes, Environment and Health Initiative (GEI) partly funded DNA extraction and statistical analyses (HG-06-033-

NCI-01 and RO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited Disease Research (U01HG004438 and NIH HHSN268200782096C), and study coordination at the GENEVA Coordination Center (U01 HG004446) for the EAGLE study and part of the PLCO. Genotyping for the remaining part of PLCO and all ATBC and CPS-II samples were supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The "Texas" study was supported by NIH grants CA55769, CA127219, R01CA133996, and CA121197. The Central European study was supported by the Institut National du Cancer (INCa) in France and the U.S. NCI (RO1 CA092039). The CARET study was supported by NIH grants R01CA78812 and U01CA63673. LUCY was partly funded by the Deutsche Forschungsgemeinschaft (DFG, BI 576/2-1; BI 576/2-2), and genotyping was funded by the Helmholtz Association in Germany. The Heidelberg sample collection was partly supported by the Deutsche Krebshilfe. The Estonian study was supported by Targeted Financing from Estonian Government (SF0180142 and ESF 6465) European Union through the European Regional Development Fund in the frame of Centre of Excellence in Genomics and 7 FP Project ECOGENE.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.08.017.

References

[1] S. Akamatsu, R. Takata, C.A. Haiman, A. Takahashi, T. Inoue, M. Kubo, et al., Common variants at 11q12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese. Nat. Genet. 44 (4) (2012) 426–429, S421http://dx.doi.org/10.1038/ng.1104.

[2] T.H. Beaty, J.C. Murray, M.L. Marazita, R.G. Munger, I. Ruczinski, J.B. Hetmanski, et al., A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat. Genet. 42 (6) (2010) 525–529, http://dx.doi.org/10.1038/ng.580.

[3] A. Bhandari, W. Gordon, D. Dizon, A.S. Hopkin, E. Gordon, Z. Yu, et al., The Grainyhead transcription factor Grhl3/Get1 suppresses miR-21 expression and tumorigenesis in skin: modulation of the miR-21 target MSH2 by RNA-binding protein DND1. Oncogene 32 (12) (2013) 1497–1507, http://dx.doi.org/10.1038/onc.2012.168.

[4] C. Bille, J.F. Winther, A. Bautz, J.C. Murray, J. Olsen, K. Christensen, Cancer risk in persons with oral cleft–a population-based study of 8,093 cases. Am. J. Epidemiol. 161 (11) (2005) 1047–1055, http://dx.doi.org/10.1093/aje/kwi132.

[5] S. Birnbaum, K.U. Ludwig, H. Reutter, S. Herms, M. Steffens, M. Rubini, et al., Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat. Genet. 41 (4) (2009) 473–477, http://dx.doi.org/10.1038/ng.333.

[6] W.J. Blot, C.A. Stiller, L.M. Wilson, Oral clefts and childhood cancer. Lancet 1 (8170) (1980) 722.

[7] C. Bonaitti-Pellie, M.L. Briard-Guillemot, J. Feingold, J. Frezal, Associated congenital malformations in retinoblastoma. Clin. Genet. 7 (1) (1975) 37–39.

[8] L.D. Botto, T. Flood, J. Little, M.N. Fluchel, S. Krikov, M.L. Feldkamp, et al., Cancer risk in children and adolescents with birth defects: a population-based cohort study. PLoS One 8 (7) (2013), e69077, http://dx.doi.org/10.1371/journal.pone.0069077 PONE-D-13-06176.

[9] L.A. Brito, G.L. Yamamoto, S. Melo, C. Malcher, S.G. Ferreira, J. Figueiredo, et al., Rare variants in the epithelial cadherin gene underlying the genetic etiology of nonsyndromic cleft lip with or without cleft palate. Hum. Mutat. 36 (11) (2015) 1029–1033, http://dx.doi.org/10.1002/humu.22827.

[10] M. Capasso, M. Devoto, C. Hou, S. Asgharzadeh, J.T. Glessner, E.F. Attiyeh, et al., Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. Nat. Genet. 41 (6) (2009) 718–723, http://dx.doi.org/10.1038/ng.374.

[11] K.Y. Chan, C.M. Wong, J.S. Kwan, J.M. Lee, K.W. Cheung, M.F. Yuen, et al., Genome-wide association study of hepatocellular carcinoma in Southern Chinese patients with chronic hepatitis B virus infection. PLoS One 6 (12) (2011), e28798, http://dx.doi.org/10.1371/journal.pone.0028798 PONE-D-11-10944.

[12] Q. Chen, H. Wang, J.B. Hetmanski, T. Zhang, I. Ruczinski, H. Schwender, et al., BMP4 was associated with NSCL/P in an Asian population. PLoS One 7 (4) (2012), e35347, http://dx.doi.org/10.1371/journal.pone.0035347 PONE-D-12-01653.

[13] K. Christensen, K. Juel, A.M. Herskind, J.C. Murray, Long term follow up study of survival associated with cleft lip and palate at birth. Br. Med. J. 328 (7453) (2004) 1405–1408 (Clin. Res. Ed).

[14] C.C. Chung, S.J. Chanock, Current status of genome-wide association studies in cancer. Hum. Genet. 130 (1) (2011) 59–78, http://dx.doi.org/10.1007/s00439-011-1030-9.

[15] W. Cozen, D. Li, T. Best, D.J. Van Den Berg, P.A. Gourraud, V.K. Cortessis, et al., A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. Blood 119 (2) (2012) 469–475, http://dx.doi.org/10.1182/blood-2011-03-343921.

[16] D. Crowther-Swanepoel, P. Broderick, M.C. Di Bernardo, S.E. Dobbins, M. Torres, M. Mansouri, et al., Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat. Genet. 42 (2) (2010) 132–136, http://dx.doi.org/10.1038/ng.510.

[17] M.G. Dunlop, S.E. Dobbins, S.M. Farrington, A.M. Jones, C. Palles, N. Whiffin, et al., Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat. Genet. 44 (7) (2012) 770–776, http://dx.doi.org/10.1038/ng.2293.

[18] D.F. Easton, K.A. Pooley, A.M. Dunning, P.D. Pharoah, D. Thompson, D.G. Ballinger, et al., Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447 (7148) (2007) 1087–1093, http://dx.doi.org/10.1038/nature05887.

[19] R.A. Eeles, Z. Kote-Jarai, A.A. Al Olama, G.G. Giles, M. Guy, G. Severi, et al., Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat. Genet. 41 (10) (2009) 1116–1121, http://dx.doi.org/10.1038/ng.450.

[20] E. Ellinghaus, M. Stanulla, G. Richter, D. Ellinghaus, G. te Kronnie, G. Cario, et al., Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. Leukemia 26 (5) (2012) 902–909, http://dx.doi.org/10.1038/leu.2011.302.

[21] V. Enciso-Mora, P. Broderick, Y. Ma, R.F. Jarrett, H. Hjalgrim, K. Hemminki, et al., A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat. Genet. 42 (12) (2010) 1126–1130, http://dx.doi.org/10.1038/ng.696.

[22] G. Evans, L. Burnell, R. Campbell, H.R. Gattamaneni, J. Birch, Congenital anomalies and genetic syndromes in 173 cases of medulloblastoma. Med. Pediatr. Oncol. 21 (6) (1993) 433–434.

[23] T. Frebourg, C. Oliveira, P. Hochain, R. Karam, S. Manouvrier, C. Graziadio, et al., Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. J. Med. Genet. 43 (2) (2006) 138–142, http://dx.doi.org/10.1136/jmg.2005.031385.

[24] M. Gazouli, M.G. Roubelakis, G.E. Theodoropoulos, J. Papailiou, A. Vaiopoulou, K.I. Pappa, et al., OCT4 spliced variant OCT4B1 is expressed in human colorectal cancer. Mol. Carcinog. 51 (2) (2012) 165–173, http://dx.doi.org/10.1002/mc.20773.

[25] M. Ghoussaini, H. Song, T. Koessler, A.A. Al Olama, Z. Kote-Jarai, K.E. Driver, et al., Multiple loci with different cancer specificities within the 8q24 gene desert. J. Natl. Cancer Inst. 100 (13) (2008) 962–966, http://dx.doi.org/10.1093/jnci/djn190.

[26] S.F. Grant, K. Wang, H. Zhang, W. Glaberson, K. Annaiah, C.E. Kim, et al., A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J. Pediatr. 155 (6) (2009) 909–913, http://dx.doi.org/10.1016/j.jpeds.2009.06.020.

[27] J. Gudmundsson, P. Sulem, D.F. Gudbjartsson, T. Blondal, A. Gylfason, B.A. Agnarsson, et al., Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat. Genet. 41 (10) (2009) 1122–1126, http://dx.doi.org/10.1038/ng.448.

[28] J. Gudmundsson, P. Sulem, A. Manolescu, L.T. Amundadottir, D. Gudbjartsson, A. Helgason, et al., Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat. Genet. 39 (5) (2007) 631–637, http://dx.doi.org/10.1038/ng1999.

[29] L.A. Hindorff, J. MacArthur, J. Morales, H.A. Junkins, P.N. Hall, A.K. Klemm, T.A. Manolio, A Catalog of Published Genome-Wide Association Studies. 2010 (Version v1.0. Available at: www.genome.gov/gwastudies. [Last accessed 12 July 2012]).

[30] R.S. Houlston, E. Webb, P. Broderick, A.M. Pittman, M.C. Di Bernardo, S. Lubbe, et al., Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat. Genet. 40 (12) (2008) 1426–1435, http://dx.doi.org/10.1038/ng.262.

[31] Z. Hu, C. Wu, Y. Shi, H. Guo, X. Zhao, Z. Yin, et al., A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat. Genet. 43 (8) (2011) 792–796, http://dx.doi.org/10.1038/ng.875.

[32] International HapMap Consortium, A haplotype map of the human genome. Nature 437 (7063) (2005) 1299–1320, http://dx.doi.org/10.1038/nature04226.

[33] I. Kluijt, E.J. Siemerink, M.G. Ausems, T.A. van Os, D. de Jong, J. Simoes-Correia, et al., CDH1-related hereditary diffuse gastric cancer syndrome: clinical variations and implications for counseling. Int. J. Cancer 131 (2) (2012) 367–376, http://dx.doi.org/10.1002/ijc.26398.

[34] V. Kumar, N. Kato, Y. Urabe, A. Takahashi, R. Muroyama, N. Hosono, et al., Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat. Genet. 43 (5) (2011) 455–458, http://dx.doi.org/10.1038/ng.809.

[35] M.T. Landi, N. Chatterjee, K. Yu, L.R. Goldin, A.M. Goldstein, M. Rotunno, et al., A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am. J. Hum. Genet. 85 (5) (2009) 679–691, http://dx.doi.org/10.1016/j.ajhg.2009.09.012.

[36] S. Li, J. Qian, Y. Yang, W. Zhao, J. Dai, J.X. Bei, et al., GWAS identifies novel susceptibility loci on 6p21.32 and 21q21.3 for hepatocellular carcinoma in chronic hepatitis B virus carriers. PLoS Genet. 8 (7) (2012), e1002791http://dx.doi.org/10.1371/journal.pgen.1002791 (PGENETICS-D-11-02735).

[37] D.E. Linn, X. Yang, F. Sun, Y. Xie, H. Chen, R. Jiang, et al., A role for OCT4 in tumor initiation of drug-resistant prostate cancer cells. Genes Cancer 1 (9) (2010) 908–916, http://dx.doi.org/10.1177/1947601910388271.

[38] H. Liu, E.J. Leslie, Z. Jia, T. Smith, M. Eshete, A. Butali, et al., Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in murine oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate. Hum. Mol. Genet. (2015)http://dx.doi.org/10.1093/hmg/ddv614.

[39] V. Lopez-Pajares, K. Qu, J. Zhang, D.E. Webster, B.C. Barajas, Z. Siprashvili, et al., A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation. Dev. Cell 32 (6) (2015) 693–706, http://dx.doi.org/10.1016/j.devcel.2015.01.028.

[40] K.U. Ludwig, E. Mangold, S. Herms, S. Nowak, H. Reutter, A. Paul, et al., Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat. Genet. 44 (9) (2012) 968–971, http://dx.doi.org/10.1038/ng.2360.
[41] E. Mangold, K.U. Ludwig, S. Birnbaum, C. Baluardo, M. Ferrian, S. Herms, et al., Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat. Genet. 42 (1) (2010) 24–26, http://dx.doi.org/10.1038/ng.506.
[42] E. Mangold, K.U. Ludwig, M.M. Nothen, Breakthroughs in the genetics of orofacial clefting. Trends Mol. Med. 17 (12) (2011) 725–733, http://dx.doi.org/10.1016/j.molmed.2011.07.007.
[43] R. Menezes, M.L. Marazita, T. Goldstein McHenry, M.E. Cooper, K. Bardi, C. Brandon, et al., AXIS inhibition protein 2, orofacial clefts and a family history of cancer. J. Am. Dent. Assoc. 140 (1) (2009) 80–84.
[44] A.C. Mertens, W. Wen, S.M. Davies, M. Steinbuch, J.D. Buckley, J.D. Potter, et al., Congenital abnormalities in children with acute leukemia: a report from the Children's Cancer Group. J. Pediatr. 133 (5) (1998) 617–623.
[45] R.W. Miller, Childhood cancer and congenital defects. A study of U.S. death certificates during the period 1960-1966. Pediatr. Dent. Restor. Dent. 3 (5) (1969) 389–397.
[46] P.A. Mossey, J. Little, Epidemiology of oral clefts: an international perspective. in: D.F. Wyszynski (Ed.), Cleft Lip and Palate: From Origin to Treatment, Oxford University Press 2002, pp. 127–144.
[47] S.A. Narod, M.M. Hawkins, C.M. Robertson, C.A. Stiller, Congenital anomalies and childhood cancer in Great Britain. Am. J. Hum. Genet. 60 (3) (1997) 474–485.
[48] M. Nishi, H. Miyake, T. Takeda, Y. Hatae, Congenital malformations and childhood cancer. Med. Pediatr. Oncol. 34 (4) (2000) 250–254.
[49] U. Peters, C.M. Hutter, L. Hsu, F.R. Schumacher, D.V. Conti, C.S. Carlson, et al., Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum. Genet. 131 (2) (2012) 217–234, http://dx.doi.org/10.1007/s00439-011-1055-0.
[50] M. Peyrard-Janvid, E.J. Leslie, Y.A. Kousa, T.L. Smith, M. Dunnwald, M. Magnusson, et al., Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am. J. Hum. Genet. 94 (1) (2014) 23–32, http://dx.doi.org/10.1016/j.ajhg.2013.11.009.
[51] S.M. Purcell, N.R. Wray, J.L. Stone, P.M. Visscher, M.C. O'Donovan, P.F. Sullivan, et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460 (7256) (2009) 748–752, http://dx.doi.org/10.1038/nature08185.
[52] S. Shete, F.J. Hosking, L.B. Robertson, S.E. Dobbins, M. Sanson, B. Malmer, et al., Genome-wide association study identifies five susceptibility loci for glioma. Nat. Genet. 41 (8) (2009) 899–904, http://dx.doi.org/10.1038/ng.407.
[53] W.L. Shih, M.W. Yu, P.J. Chen, S.H. Yeh, M.T. Lo, H.C. Chang, et al., Localization of a susceptibility locus for hepatocellular carcinoma to chromosome 4q in a hepatitis B hyperendemic area. Oncogene 25 (22) (2006) 3219–3224, http://dx.doi.org/10.1038/sj.onc.1209345.
[54] C.F. Skibola, P.M. Bracci, E. Halperin, L. Conde, D.W. Craig, L. Agana, et al., Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat. Genet. 41 (8) (2009) 873–875, http://dx.doi.org/10.1038/ng.419.
[55] S.N. Stacey, P. Sulem, G. Masson, S.A. Gudjonsson, G. Thorleifsson, M. Jakobsdottir, et al., New common variants affecting susceptibility to basal cell carcinoma. Nat. Genet. 41 (8) (2009) 909–914, http://dx.doi.org/10.1038/ng.412.
[56] E.F. Steinwachs, C. Amos, D. Johnston, J. Mulliken, S. Stal, J.T. Hecht, Nonsyndromic cleft lip and palate is not associated with cancer or other birth defects. Am. J. Med. Genet. 90 (1) (2000) 17–24.
[57] J. Sterneckert, S. Hoing, H.R. Scholer, Concise review: Oct4 and more: the reprogramming expressway. Stem Cells 30 (1) (2012) 15–21, http://dx.doi.org/10.1002/stem.765.
[58] J. Suazo, J.C. Tapia, J.L. Santos, V.G. Castro, A. Colombo, R. Blanco, Risk variants in BMP4 promoters for nonsyndromic cleft lip/palate in a Chilean population. BMC Med. Genet. 12 (2011) 163, http://dx.doi.org/10.1186/1471-2350-12-163.
[59] S. Suzuki, M.L. Marazita, M.E. Cooper, N. Miwa, A. Hing, A. Jugessur, et al., Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am. J. Hum. Genet. 84 (3) (2009) 406–411, http://dx.doi.org/10.1016/j.ajhg.2009.02.002.
[60] A. Tenesa, S.M. Farrington, J.G. Prendergast, M.E. Porteous, M. Walker, N. Haq, et al., Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat. Genet. 40 (5) (2008) 631–637, http://dx.doi.org/10.1038/ng.133.
[61] G. Thomas, K.B. Jacobs, M. Yeager, P. Kraft, S. Wacholder, N. Orr, et al., Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet. 40 (3) (2008) 310–315, http://dx.doi.org/10.1038/ng.91.
[62] I. Tomlinson, E. Webb, L. Carvajal-Carmona, P. Broderick, Z. Kemp, S. Spain, et al., A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat. Genet. 39 (8) (2007) 984–988, http://dx.doi.org/10.1038/ng2085.
[63] C. Turnbull, S. Ahmed, J. Morrison, D. Pernet, A. Renwick, M. Maranian, et al., Genome-wide association study identifies five new breast cancer susceptibility loci. Nat. Genet. 42 (6) (2010) 504–507, http://dx.doi.org/10.1038/ng.586.
[64] V.V. Uslu, M. Petretich, S. Ruf, K. Langenfeld, N.A. Fonseca, J.C. Marioni, et al., Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat. Genet. 46 (7) (2014) 753–758, http://dx.doi.org/10.1038/ng.2971.
[65] Y. Wang, P. Broderick, E. Webb, X. Wu, J. Vijayakrishnan, A. Matakidou, et al., Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat. Genet. 40 (12) (2008) 1407–1409, http://dx.doi.org/10.1038/ng.273.
[66] Z. Wang, E. Oron, B. Nelson, S. Razis, N. Ivanova, Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell 10 (4) (2012) 440–454, http://dx.doi.org/10.1016/j.stem.2012.02.016.
[67] S. Yang, J. Zheng, Y. Ma, H. Zhu, T. Xu, K. Dong, et al., Oct4 and Sox2 are overexpressed in human neuroblastoma and inhibited by chemotherapy. Oncol. Rep. 28 (1) (2012) 186–192, http://dx.doi.org/10.3892/or.2012.1765.
[68] M. Yeager, N. Orr, R.B. Hayes, K.B. Jacobs, P. Kraft, S. Wacholder, et al., Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat. Genet. 39 (5) (2007) 645–649, http://dx.doi.org/10.1038/ng2022.
[69] M. Zack, H.O. Adami, A. Ericson, Maternal and perinatal risk factors for childhood leukemia. Cancer Res. 51 (14) (1991) 3696–3701.
[70] H. Zhang, Y. Zhai, Z. Hu, C. Wu, J. Qian, W. Jia, et al., Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. Nat. Genet. 42 (9) (2010) 755–758, http://dx.doi.org/10.1038/ng.638.
[71] Z. Zhang, Y. Song, X. Zhao, X. Zhang, C. Fermin, Y. Chen, Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. Development 129 (17) (2002) 4135–4146.
[72] J.L. Zhu, O. Basso, H. Hasle, J.F. Winther, J.H. Olsen, J. Olsen, Do parents of children with congenital malformations have a higher cancer risk? A nationwide study in Denmark. Br. J. Cancer 87 (5) (2002) 524–528, http://dx.doi.org/10.1038/sj.bjc.6600488.

3. Danksagung

Zunächst möchte ich Herrn Prof. Dr. med. Markus Nöthen für seine Betreuung danken und dafür, dass er mir die Möglichkeit gegeben hat, in der Arbeitsgruppe „Genetik orofazialer Spalten“ am Institut für Humangenetik mitzuarbeiten.
Mein ganz besonderer Dank geht an PD Dr. med. Elisabeth Mangold aus der Arbeitsgruppe „Genetik orofazialer Spalten“ für die Anleitung und Führung meiner Arbeit, die Unterstützung und Betreuung, die vielen Ideen und die angenehme und gute Zusammenarbeit seit Beginn an.
Weiterhin möchte ich Frau Dr. rer. nat. Kerstin Ludwig aus der Arbeitsgruppe „Genetik orofazialer Spalten“ für ihre Unterstützung, die vielen konstruktiven Anregungen und Ideen, Korrekturen und Hinweise danken. Ich danke zudem Herrn PD Dr. rer. nat. Michael Knapp vom Institut für Medizinische Biometrie, Informatik und Epidemiologie für die statistische Auswertung und Beratung. Auch Herr Dr. med. Manuel Mattheisen und Frau Dr. rer. nat. Anne Böhmer haben mir sehr bei der Erstellung meiner Arbeit geholfen.
Ich danke den zahlreichen Mitarbeitern für die Erstellung der NSCL/P- Datensätze und dass ich diese nutzen durfte. Mein Dank geht auch an die Patienten und Familien für Ihre Teilnahme an den vorausgegangenen Studien und Analysen.
Ich danke außerdem den Arbeitsgruppen von Christine F. Skibola, Jane C. Figueiredo, Fay Julie Hosking, Eva Ellinghaus, Maria Teresa Landi, Hongxia Ma, Hidewaki Nakagawa, Jong-Won Kim, Jiali Han, Ping Yang, Victor Enciso, Susan L. Slager, Peter Kraft und deCODE Genetics für die Bereitstellung der Daten aus ihren GWAS zu Krebserkrankungen.
Schließlich möchte ich meiner Familie für die Unterstützung während der Erstellung meiner Arbeit danken.